



**Уральский
федеральный
университет**

имени первого Президента
России Б.Н.Ельцина

**Высшая школа
экономики
и менеджмента**

МЕТОДЫ И МОДЕЛИ ПРИКЛАДНЫХ ИССЛЕДОВАНИЙ

Учебное пособие

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИМЕНИ ПЕРВОГО ПРЕЗИДЕНТА РОССИИ Б. Н. ЕЛЬЦИНА

МЕТОДЫ И МОДЕЛИ ПРИКЛАДНЫХ ИССЛЕДОВАНИЙ

Учебное пособие

Рекомендовано
методическим советом Уральского федерального университета
в качестве учебного пособия для студентов вуза,
обучающихся по направлению подготовки
38.03.01, 38.04.01 «Экономика»

Екатеринбург
Издательство Уральского университета
2019

УДК 330.43(075.8)
М545

Авторы:

О. С. Мариев, И. В. Савин, А. А. Пушкарев,
И. С. Шорохова, А. Н. Жуков, Н. В. Кисляк

Под общей редакцией
О. С. Мариева

Рецензенты:

сектор инфраструктурного развития
и экономико-математических методов Института экономики УрО РАН
(заведующий сектором кандидат экономических наук, доцент С. Н. Котлярова);
Н. М. Сурнина, доктор экономических наук, профессор,
заведующий кафедрой информационных технологий и статистики
Уральского государственного экономического университета

Методы и модели прикладных исследований : учеб. пособие /
М545 [О. С. Мариев, И. В. Савин, А. А. Пушкарев, И. С. Шорохова, А. Н. Жуков, Н. В. Кисляк ; под общ. ред. О. С. Мариева] ; М-во науки и высш. образования Рос. Федерации, Урал. федер. ун-т. — Екатеринбург : Изд-во Урал. ун-та, 2019. — 232 с.

ISBN 978-5-7996-2664-8

В учебном пособии дается обзор современных эмпирических методов и моделей, которые могут быть использованы в прикладных исследованиях экономических, социальных и иных процессов и явлений. В первом разделе пособия рассматриваются основные эконометрические модели, особенности их применения, анализируются временные ряды и панельные данные. Второй раздел посвящен методам Монте-Карло и их применению в современных исследованиях, а также основам агента-базированного моделирования и применению этих моделей.

Для студентов и аспирантов экономических специальностей.

УДК 330.43(075.8)

ОГЛАВЛЕНИЕ

Предисловие.....	5
------------------	---

РАЗДЕЛ I. ЭКОНОМЕТРИКА

Глава 1. Введение в эконометрику.....	8
Корреляционная связь.....	8
Сущность, этапы эконометрического моделирования, основные модели и их особенности.....	17
Эконометрические методы оценки параметров модели.....	23
Тестирование статистических гипотез.....	39
Глава 2. Нарушение основных предположений МНК-оценки.....	46
Гетероскедастичность.....	46
Автокорреляция.....	52
Мультиколлинеарность.....	55
Эндогенность.....	59
Глава 3. Системы регрессионных уравнений.....	82
Внешне не связанные регрессии.....	82
Модели одновременных уравнений.....	83
Глава 4. Временные ряды и панельные данные.....	92
Введение во временные ряды.....	92
AR-, MA-, ARMA-модели.....	102
Автокорреляция и частичная автокорреляция. Функции ACF и PACF.....	106

Введение в панельные данные.....	108
Виды оценок, применяемых для панельных данных.....	112
Глава 5. Модели бинарного выбора.....	122
Линейная модель вероятности.....	122
Logit- и probit-модели.....	124
Глава 6. Мультиномиальные и упорядоченные модели.....	129
Мультиномиальные модели.....	129
Упорядоченные модели.....	131
Цензурированные и усеченные модели регрессии.....	135

РАЗДЕЛ II. МЕТОДЫ МОНТЕ-КАРЛО

Глава 1. Введение в методы Монте-Карло.....	139
Методы Монте-Карло.....	140
Вероятность.....	147
Генерация случайных чисел.....	150
Несколько применений методов Монте-Карло.....	156
Глава 2. Сэмплирование данных.....	159
Сэмплирование из стандартных распределений.....	159
Сэмплирование из нестандартных распределений.....	162
Глава 3. Эвристические методы оптимизации.....	170
Обзор эвристических методов.....	172
Сходимость эвристики.....	180
Приложение к дискретным пространствам поиска.....	188
Приложение к непрерывным пространствам поиска.....	190
Глава 4. Введение в агента-базированное моделирование.....	193
Обзор агента-базированных моделей.....	193
Примеры использования АВМ.....	203
Заключение.....	218
Библиографические ссылки.....	221
Список рекомендуемой литературы.....	223

ПРЕДИСЛОВИЕ

Данное учебное пособие посвящено промежуточному уровню эконометрического анализа. Целевая аудитория книги — студенты первого курса магистратуры, которые имеют базовые знания по математическому анализу, статистике, алгебре и теории вероятности. Эконометрический анализ является очень важным в современных экономических исследованиях и требует глубокого понимания различных процессов и методов оценки. Эконометрика применяется во всех прикладных областях экономики для тестирования экономических теорий, формирования рекомендаций, используемых при выработке государственных и управленческих решений, а также прогнозирования экономических временных рядов. Иногда эконометрическая модель вытекает из формальной экономической модели, но обычно эконометрические модели основываются на неформальных экономических рассуждениях и интуиции. Целью любого эконометрического анализа является оценка параметров модели и проверка гипотез об этих параметрах; значения и знаки параметров определяют значимость экономической теории и эффекты от определенной политики. В книге рассматриваются все основные инструменты и концепции, необходимые для эмпирического анализа.

В пособии широко используется математический язык как одно из средств объяснения, поэтому студент должен обладать определенным уровнем математических навыков, включая матричную арифметику и понимание интегрального исчисления. Кроме того, в данной книге приведены алгоритмы работы с компьютерными программами, использующими инструменты MATLAB в качестве иллюстраций, для их понимания опыт программирования будет полезен, но не критически необходим, так как все примеры комментируются и поясняются. «Компьютерное» мышление более важно для понимания. Понять материал поможет знание статистики, теории вероятностей и эконометрики.

Первый раздел посвящен основам эконометрики — анализу простой линейной регрессии, тестированию гипотез, основным положениям МНК-оценки и их главным нарушениям, а именно проблемам гетероскедастичности, автокорреляции, мультиколлинеарности и эндогенности. Даются основы анализа с использованием различных структур и типов данных: системы уравнений, временных рядов, панельных данных, моделей бинарного выбора, мультиномиальных и упорядоченных моделей.

Второй раздел состоит из четырех глав, в которых рассматриваются различные компоненты методов Монте-Карло и их реализация.

В главе 1 дается понятие методов Монте-Карло: как они появились и развивались, какие инструменты применяются для их использования. Кроме того, пересматриваются понятия вероятности и случайности, необходимые для понимания данного подхода. Объясняется, как работают различные генераторы случайных чисел, и раскрываются их отличительные особенности. В конце главы приводятся несколько основных примеров того, как можно использовать методы Монте-Карло на практике.

Во второй главе объясняется, что такое выборка, почему выборки используются на практике; демонстрируется, как можно использовать стандартные распределения, приводятся список команд для MATLAB и примеры кода. Далее следует объяснение более сложной процедуры выборки из нестандартных распределений с использованием методов Монте-Карло.

В третьей главе рассматривается концепция эвристического подхода к оптимизации, дается объяснение различных типов эвристических алгоритмов оптимизации и показывается, как они должны быть построены и реализованы. В качестве иллюстраций представлено несколько применений эвристической оптимизации для дискретных и непрерывных пространств поиска.

В четвертой главе объясняется современная концепция агентного моделирования — общие моменты его идеи, составляющие, почему этот вид моделирования может быть более эффективным, чем классические модели, и основные отличительные черты. Кроме того, здесь представлено несколько примеров агентных моделей (городские модели, модели динамики мнений, модели цепочек поставок и промышленных сетей), охватывающих все основные сферы применения моделирования.

В качестве примера в данном пособии используется синтаксис программного обеспечения MATLAB для программирования иллюстраций алгоритмов и создания информативных графиков и визуализаций результатов. При выборе языка программирования мы исходили в основном из простоты его понимания и возможностей использования для решения задач, так как он предоставляет широкий спектр математических и графических инструментов. Приведенные в книге коды достаточно просты и могут быть реализованы в среде MATLAB (или ее бесплатной альтернативе — Octave). Однако при решении сложных задач, требующих реализации методов Монте-Карло, исследователи могут рассмотреть возможность использования более богатых языков программирования, таких как Java, Python или C++.

В пособии также представлены различные алгоритмы генерации случайных чисел и их практическое использование в методах Монте-Карло.

Данное учебное пособие может быть полезно при изучении курса по эконометрике.

Раздел I

ЭКОНОМЕТРИКА

ГЛАВА 1. ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ

Корреляционная связь

Изучая экономические явления, исследователи сталкиваются с зависимостями как между количественными, так и между качественными признаками, поэтому основной задачей статистики является обнаружение таких зависимостей и определение их количественной характеристики. Связь признаков можно обнаружить только при массовом наблюдении. Выявленная таким образом связь называется статистической. Частный случай статистической связи — связь корреляционная.

Корреляционная связь — это связь, проявляющаяся при большом числе наблюдений в виде определенной зависимости между средним значением результативного признака и признаками-факторами. Если рассматривается связь средней величины результативного показателя y с одним признаком-фактором x , корреляция называется парной, а если факторных признаков два и более (x_1, x_2, \dots, x_m) , корреляция называется множественной.

По характеру изменений x и y в парной корреляции различают прямую и обратную связь. При прямой связи значения обоих признаков изменяются в одном направлении, т. е. с увеличением

(уменьшением) значений x увеличиваются (уменьшаются) и значения y . При обратной связи значения факторного и результативного признаков изменяются в разных направлениях.

Статистическое изучение корреляционной связи сводится к решению трех задач: выявить наличие или отсутствие такой связи, измерить тесноту связи между исследуемыми признаками и определить математическую модель такой взаимосвязи.

Для установления наличия и характера корреляционной связи между признаками в статистике используются следующие методы выявления корреляционной связи: метод рассмотрения параллельных данных, коэффициент корреляции знаков Фехнера, графический метод, метод аналитических группировок и метод корреляционных таблиц.

Метод рассмотрения параллельных данных (значений x и y в каждой из n единиц) состоит в том, что единицы наблюдения располагают по возрастанию значений факторного признака x и затем сравнивают с ним визуально поведение результативного признака y .

Коэффициент корреляции знаков Фехнера — простейший показатель тесноты связи, основанный на сравнении поведения отклонений индивидуальных значений признака x и результативного признака y от своих средних величин. При этом во внимание принимаются не величины отклонений, а их знаки («+» или «-»). Определив знаки отклонений от средней величины в каждом ряду, рассматривают все пары знаков и подсчитывают число их совпадений (n_a) и несовпадений (n_b). Тогда коэффициент Фехнера рассчитывается как отношение разности чисел пар совпадений и несовпадений знаков к их сумме, т. е. к общему числу наблюдаемых единиц:

$$K_{\Phi} = \frac{\sum n_a - \sum n_b}{\sum n_a + \sum n_b}.$$

Очевидно, что если знаки всех отклонений по каждому признаку совпадут, то $K_{\Phi} = 1$, что характеризует наличие прямой связи. Если все знаки не совпадут, то $K_{\Phi} = -1$ (обратная связь). Если же $\sum n_a = \sum n_b$, то $K_{\Phi} = 0$.

Итак, как и любой показатель тесноты связи, коэффициент Фехнера может принимать значения от 0 до ± 1 . Однако если $K_{\phi} = 1$, то это ни в коей мере нельзя воспринимать как свидетельство функциональной зависимости между x и y .

Графический метод — это графическое изображение корреляционной зависимости, когда каждую пару взаимосвязанных значений x и y изображают в виде точки на плоскости с координатами x и y в прямоугольной системе координат. Совокупность полученных точек представляет собой корреляционное поле, а соединяя последовательно нанесенные точки отрезками, получают ломаную линию, именуемую эмпирической линией регрессии. Визуально анализируя график, можно предположить характер зависимости между признаками x и y .

Метод аналитических группировок используется при большом числе наблюдений для выявления корреляционной связи между двумя количественными признаками. Чтобы выявить наличие корреляционной связи между двумя признаками, проводится группировка единиц совокупности по факторному признаку x и для каждой выделенной группы рассчитывается среднее значение результативного признака y . Если результативный признак y зависит от факторного признака x , то в изменении среднего значения будет прослеживаться определенная закономерность.

Метод корреляционных таблиц предполагает комбинационное распределение в таблице единиц совокупности по двум количественным признакам. Такая таблица строится по типу шахматной, т. е. в подлежащем (строках) таблицы указаны группы по факторному признаку x , в сказуемом (столбцах) — по результативному признаку y (или наоборот), а в клетках таблицы на пересечении x и y отмечено число случаев совпадения каждого значения x с соответствующим значением y . О наличии и направлении связи можно судить по внешнему виду таблицы, т. е. по расположению в ней частот. Если частоты расположены в клетках таблицы беспорядочно, то это чаще всего свидетельствует об отсутствии связи между группировочными признаками (или об их незначительной взаимозависимости); если частоты тяготеют ближе к одной из диагоналей и центру таблицы, образуя своего рода эллипс, то это почти всегда

свидетельствует о наличии зависимости между признаками x и y , близкой к линейной. Расположение по диагонали из верхнего левого угла в нижний правый свидетельствует о прямой линейной связи, а из нижнего левого угла в верхний правый — об обратной.

Метод корреляционных таблиц может быть использован и для таблиц сопряженности различной размерности. Простейшая размерность — 2×2 (таблица «четырёх полей»), когда по альтернативному признаку («да» — «нет», «мужской пол» — «женский» и т. д.) выделяются 2 группы признаков.

В корреляционном анализе важно не только определить наличие связи между исследуемыми показателями, но и измерить тесноту данной связи. Выделяют две группы показателей степени тесноты связи: между двумя и более качественными показателями и между двумя и более количественными показателями.

Показатели степени тесноты связи между качественными показателями. В таблицах сопряженности 2×2 для измерения этих показателей используют коэффициент ассоциации Юла и коэффициент контингенции Пирсона; в таблицах большей размерности (два и более признака) определяют коэффициенты взаимной сопряженности Пирсона и Чупрова, рассчитываемые на основе показателя χ^2 .

Формулы для расчета названных коэффициентов:

— коэффициент ассоциации Юла:

$$K_a = \frac{ad - bc}{ad + bc},$$

— коэффициент контингенции Пирсона:

$$K_k = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

(связь считается достаточно значительной и подтвержденной, если $|K_a| > 0,5$ или $|K_k| > 0,3$);

— коэффициент взаимной сопряженности Пирсона:

$$K_{\Pi} = \sqrt{\frac{\chi^2}{\chi^2 + n}},$$

— коэффициент взаимной сопряженности Чупрова:

$$K_{\text{ч}} = \sqrt{\frac{\chi^2}{n\sqrt{(k_1 - 1)(k_2 - 1)}}}.$$

Рассчитывать коэффициент Чупрова для таблицы четырех полей не рекомендуется, так как при числе степеней свободы $\nu = (2 - 1)(2 - 1) = 1$ он будет больше коэффициента Пирсона. Для таблиц же большей размерности всегда $K_{\text{ч}} < K_{\text{п}}$.

Для оценки наличия или отсутствия зависимости между признаками в клетках таблицы сопряженности используют критерий Пирсона χ^2 . Он позволяет судить о случайности (или неслучайности) распределения в таблицах взаимной сопряженности и, следовательно, об отсутствии или наличии зависимости между признаками группировки в таблице. Чтобы воспользоваться критерием Пирсона χ^2 , в таблице взаимной сопряженности наряду с эмпирическими частотами записывают теоретические частоты, рассчитываемые исходя из предположения, что распределение внутри таблицы случайно и, следовательно, зависимость между признаками группировки отсутствует. То есть считается, что распределение частот в каждой строке (столбце) таблицы пропорционально распределению частот в итоговой строке (столбце). Поэтому теоретические частоты по строкам (столбцам) рассчитывают пропорционально распределению единиц в итоговой строке (столбце).

Рассчитанное (фактическое) значение χ^2 сопоставляют с табличным, определяемым по таблице для заданного уровня значимости α и числа степеней свободы, где k_1 и k_2 — число групп по одному и второму признакам группировки (число строк и число столбцов в таблице). Если $\chi^2 \neq 0$, то значит между рассматриваемыми показателями существует зависимость. При независимости признаков в таблице сопряженности частоты теоретического и эмпирического распределений совпадают, а значит, $\chi^2 = 0$.

Показатели степени тесноты связи между двумя и более количественными показателями. К этой группе показателей относят: линейный коэффициент корреляции, коэффициенты корреляции рангов Спирмена и Кендалла, коэффициент конкордации, коэффи-

циент эластичности. Также степень тесноты связи между количественными показателями можно оценить с помощью коэффициента корреляции знаков Фехнера.

Линейный коэффициент корреляции (r) — самый популярный измеритель тесноты линейной связи между двумя количественными признаками x и y . Он основан на предположении, что при полной независимости признаков x и y отклонения значений факторного признака от средней носят случайный характер и должны случайно сочетаться с различными отклонениями. При наличии значительного перевеса совпадений или несовпадений таких отклонений делается предположение о наличии связи между x и y .

В отличие от K_Φ в линейном коэффициенте корреляции учитываются не только знаки отклонений от средних величин, но и значения самих отклонений, выраженные для сопоставимости в единицах среднего квадратического отклонения t :

$$t_x = \frac{x - \bar{x}}{\sigma_x} \quad \text{и} \quad t_y = \frac{y - \bar{y}}{\sigma_y}.$$

Линейный коэффициент корреляции r представляет собой среднюю величину из произведений нормированных отклонений для x и y :

$$r = \frac{\sum \left(\frac{x - \bar{x}}{\sigma_x} \right) \left(\frac{y - \bar{y}}{\sigma_y} \right)}{n} = \frac{\sum t_x t_y}{n} \quad \text{или} \quad r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y}.$$

Линейный коэффициент корреляции может принимать значения от -1 до $+1$, причем знак определяется в ходе решения. Например, если r по формуле будет положительным, то существует прямая зависимость между x и y ; в противном случае ($r < 0$) имеет место обратная связь. Если $r = 0$, то это означает отсутствие линейной зависимости между x и y , а при $r = 1$ наблюдается полная зависимость (функциональная) между x и y . Следовательно, всякое промежуточное значение r от 0 до 1 характеризует степень приближения корреляционной связи между x и y к функциональной.

Существует эмпирическое правило (шкала Чеддока) для оценки тесноты связи.

Таким образом, коэффициент корреляции при линейной зависимости служит и мерой тесноты связи, и показателем, характеризующим степень приближения корреляционной зависимости между x и y к линейной. Поэтому близость значения r к 0 в одних случаях может означать отсутствие связи между x и y , а в других может свидетельствовать о том, что зависимость нелинейная.

Линейный коэффициент корреляции необходимо проверять на значимость (существенность). Интерпретируя значение линейного коэффициента корреляции, следует иметь в виду, что он рассчитан для ограниченного числа наблюдений и подвержен случайным колебаниям, как и сами значения x и y , на основе которых он рассчитан. Другими словами, как любой выборочный показатель, он содержит случайную ошибку и не всегда однозначно отражает действительно реальную связь между изучаемыми показателями. Для того чтобы оценить существенность (значимость) самого линейного коэффициента корреляции r и соответственно реальность измеряемой связи между x и y , необходимо рассчитать среднюю квадратическую ошибку коэффициента корреляции σ_r . Оценка существенности (значимости) r основана на сопоставлении значения r с его средней квадратической ошибкой: $|r|/\sigma_r$.

Существуют некоторые особенности расчета σ_r в зависимости от числа наблюдений n (объема выборки).

Если число наблюдений достаточно велико ($n > 30$), то σ_r рассчитывается по формуле

$$\sigma_r = \frac{1-r^2}{\sqrt{n}}.$$

Обычно если $|r|/\sigma_r > 3$, то r считается значимым (существенным), а связь — реальной.

Если число наблюдений небольшое ($n < 30$), то σ_r рассчитывается по формуле

$$\sigma_r = \frac{\sqrt{1-r^2}}{\sqrt{n-2}},$$

а значимость r проверяется на основе t -критерия Стьюдента, для чего определяется расчетное значение критерия по формуле

$$t_{\text{расч}} = \frac{|r|}{\sigma_r} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.$$

Расчетное значение $t_{\text{расч}}$ сопоставляется с табличным значением $t_{\text{табл}}$, которое находят по таблице распределения t -критерия Стьюдента при уровне значимости $\alpha = 1 - \beta$ и числе степеней свободы $df = n - 2$. Если $t_{\text{расч}} > t_{\text{табл}}$, то r считается значимым, а связь между x и y — реальной. В противном случае ($t_{\text{расч}} < t_{\text{табл}}$) считается, что связь между x и y отсутствует и значение r , отличное от нуля, получено случайно.

Измерить тесноту связи между коррелируемыми величинами — значит определить, насколько вариация результативного признака обусловлена вариацией факторного (факторных) признака (признаков). Ранее были рассмотрены показатели, с помощью которых можно выявить наличие корреляционной связи между двумя признаками x и y и измерить тесноту этой связи.

Наряду с ними существует универсальный показатель — корреляционное отношение (или коэффициент корреляции Пирсона), применимое ко всем случаям корреляционной зависимости независимо от формы этой связи. Следует различать эмпирическое и теоретическое корреляционное отношение.

Эмпирическое корреляционное отношение (η) рассчитывается на основе правила сложения дисперсий, как корень квадратный из отношения межгрупповой дисперсии к общей дисперсии:

$$\eta = \sqrt{\frac{\delta^2}{D_0}}.$$

Теоретическое корреляционное отношение ($\eta_{\text{теор}}$) определяется как корень квадратный из теоретического коэффициента детерминации:

$$\eta_{\text{теор}} = \sqrt{\eta_{\text{теор}}^2}.$$

Теоретический коэффициент детерминации показывает, какую долю в общей дисперсии результативного признака занимает дисперсия, выражающая влияние вариации фактора x на вариацию y , и рассчитывается по формуле:

$$\eta_{\text{теор}}^2 = \frac{\delta^2}{\sigma_y^2} = \frac{\sum (\hat{y}_x - \bar{y})^2}{\sum (y_i - \bar{y})^2},$$

где $\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}$ — дисперсия эмпирического ряда;

$\delta^2 = \frac{\sum (\hat{y}_x - \bar{y})^2}{n}$ — дисперсия теоретического ряда.

Теоретическое корреляционное отношение может находиться в пределах от 0 до 1; чем ближе его значение к 1, тем теснее связь между вариацией x и y . Для оценки тесноты связи обычно применяется шкала Чеддока.

Корреляционное отношение применимо как для парной, так и для множественной корреляции независимо от формы ее связи. В этом смысле корреляционное отношение можно назвать универсальным показателем тесноты связи.

С помощью рассмотренных методов можно доказать наличие связи, установить ее направление и степень тесноты, но нельзя определить значимость изучаемого фактора среди других факторов. Чтобы выяснить последнее, надо установить теоретическую форму связи между изучаемыми признаками, т. е. необходимо найти тип функции, позволяющий наиболее адекватно отразить зависимость между признаками x и y , называемый уравнением регрессии, о котором пойдет речь далее.

Сущность, этапы эконометрического моделирования, основные модели и их особенности

Большая часть современного экономического анализа посвящена проблеме, как различные аспекты нашей жизни связаны друг с другом и как найти оптимальную модель, описывающую некоторый экономический факт. Например, какие факторы влияют на уровень заработной платы или как уровень образования влияет на уровень заработной платы в будущем. С использованием реальных данных и экономической теории эконометрика позволяет выявить эти связи.

Термин «эконометрика» был введен австро-венгерским бухгалтером в 1910 г. Он считал, что применение математических методов в данных бухгалтерского учета может дать новый обзор результатов. Но его концепция была забыта, в то время как термин «эконометрика» превратился в название новой дисциплины в экономической науке.

Для начала необходимо дать определение эконометрике. *Эконометрика* — это дисциплина, которая объединяет теоретические методы и модели. Это своего рода применение статистических и математических методов в направлении тестирования экономических гипотез, нахождения разных отношений в реальном мире и оценке государственной политики.

Первым шагом для достижения поставленных целей в эконометрическом анализе должно стать построение эконометрической модели. Модель является функцией от ряда переменных:

$$Y = f(x_1, x_2, \dots, x_n),$$

где Y — зависимая переменная; x_1, x_2, \dots, x_n — независимые (объясняющие) переменные.

Пример:

$$wage = f(education, experience, age, \dots).$$

Как правило, выделяются следующие *этапы эконометрического моделирования*:

1. Экономическая интуиция. На первом этапе необходимо проанализировать данные и общую экономическую ситуацию и за-

метить, что некоторые переменные коррелируют друг с другом. Например, рыночный спрос зависит от цены, издержек, цен на товары-заменители, спроса на комплиментарные товары и т. д. После этого нужно выбрать зависимую переменную и объясняющие (или независимые) переменные. Если модель состоит из достаточно большого числа независимых переменных (которые являются значимыми в вашей модели), то она обладает высокой объяснительной силой.

2. Следующий этап эконометрического моделирования предполагает сбор данных и выбор типа данных. Существуют следующие типы данных в эконометрике:

а) *пространственные данные* — это случайная выборка некоторых объектов (переменных), собранных в определенный момент времени:

$$(x_{1i}, x_{2i}, \dots, x_{ki}, y_i), i = 1, \dots, N.$$

Например, если мы хотим измерить текущий уровень дохода в какой-либо стране, то мы можем опросить выборку из 10 тыс. человек, отобранных случайным образом. Мы можем узнать об их месячной заработной плате (примечание: обязательно берите данные по заработной плате за один и тот же месяц) и после этого посчитать, какой процент выборки имеет заработную плату ниже средней (или что-либо другое, интересное нам). Согласно примеру мы видим, что пространственные данные представляют только текущее состояние, мы не знаем, будет ли процент людей с низкой зарплатой становиться больше или меньше;

б) *временные ряды* — это данные, которые были собраны в течение определенного периода времени (они могут содержать одну или более переменных, но в течение конкретного периода времени):

$$Y_1, Y_2, \dots, Y_T, T — \text{охват выборки.}$$

Так, для того чтобы измерить, как заработная плата агента меняется с течением времени, мы должны собрать данные о его заработной плате за определенный период времени, например, сведения о заработной плате за последние 5 лет. После этого мы можем рассчитать динамику ежемесячной индивидуальной заработной платы;

в) *панельные данные* — данные, которые содержат наблюдения некоторых переменных из разных экономических объектов за определенные последовательные периоды времени (другими словами, панельные данные имеют размерность и временных рядов, и пространственных данных).

Пусть X_{jit} — j -я переменная из объекта i в момент времени t .

$$X_{jit} = \begin{pmatrix} x_{1it} \\ x_{2it} \\ \dots \\ x_{kit} \end{pmatrix},$$

где $i = 1, \dots, N$ и $t = 1, \dots, T$.

Панельные данные объединяют пространственные данные и временные ряды и показывают множество объектов и как они меняются с течением времени.

Например, предположим, что мы имеем данные о заработной плате, образовании и опыте работы для ряда лиц за последние пять лет или мы хотим сравнить торговые потоки из ряда стран с другой совокупностью стран за 10-летний период.

3. На третьем этапе нам нужно выбрать спецификацию модели:

- а) формулировка цели моделирования;
- б) определение экзогенных и эндогенных переменных в модели;
- в) определение отношений зависимости между переменными;
- г) установка предварительного ограничения для случайной составляющей;
- д) формулировка предварительного ограничения для коэффициентов.

Типы эконометрических моделей:

а) модели временных рядов: объяснение переменной, которая изменяется с течением времени, на основе ее предыдущих значений (например, модели тренда и сезонности);

б) регрессионные модели с одним уравнением: зависимая переменная является функцией от независимых (объясняющих) переменных и случайной ошибки. Модель может быть линейной или нелинейной, это зависит от вида функции модели.

Пример:

Функция выпуска:

$$Y = \beta_0 + \beta_1 K + \beta_2 L + \varepsilon$$

или

$$Y = AL^\alpha K^\beta;$$

в) одновременные уравнения: поведение агента описывается системой уравнений:

$$\begin{cases} Q_t^D = \alpha_1 + \beta_1 p_t + \varepsilon_t^D \\ Q_t^S = \alpha_2 + \beta_2 p_t + \varepsilon_t^S \\ Q_t^S = Q_t^D \end{cases}$$

Эта система регрессионных уравнений состоит из уравнения спроса, уравнения предложения и уравнения рыночного равновесия (равенство уравнений спроса и предложения).

4. Следующий этап эконометрического моделирования — идентификация модели. Она предполагает статистический анализ модели, в первую очередь статистические оценки параметров модели. Главной частью анализа является выбор метода оценки и тестирования различных гипотез.

5. Верификация модели: сопоставление реальных данных и полученных результатов, проверка корректности результатов, которые показывает модель. В противном случае необходимо изменять модель в соответствии с экономической интуицией и собранными данными.

Функция регрессии может быть представлена по-разному. Чтобы выбрать подходящую, следует учитывать: экономическую теорию, опыт, интуицию исследователя, эмпирический анализ данных. В парном случае набор данных представлен как пара множеств переменных:

$$(x_i, y_i), i = 1, 2, \dots, N.$$

Каждая пара чисел может быть нанесена на график, как показано на графической иллюстрации (рис. 1):

Этот график также называется «облако наблюдения», «область корреляции», «точечная диаграмма». Диаграмма помогает определить вид функции регрессии.

Виды функции регрессии:

1. Линейная функция регрессии (рис. 2, 3):

$$y = \alpha + \beta x + \varepsilon.$$

2. Квадратичная функция регрессии (рис. 4):

$$y = \alpha + \beta x + \gamma x^2 + \varepsilon.$$

В этой функции существует одна переменная (x), но две независимые переменные (x и x^2).

3. Показательная функция регрессии (рис. 5):

$$y = \alpha x^{\beta} \varepsilon.$$

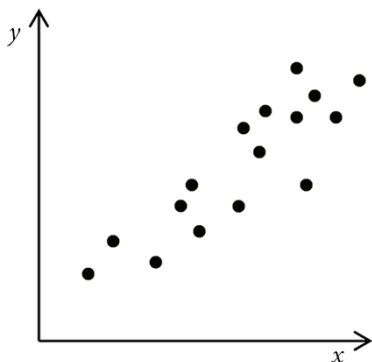


Рис. 1. Диаграмма рассеяния

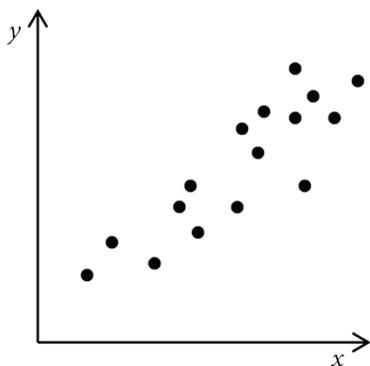


Рис. 2. Диаграмма рассеяния линейной регрессионной модели (тип 1)

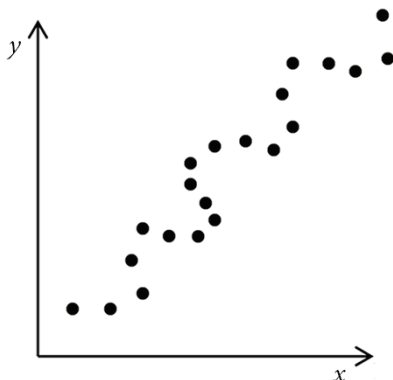


Рис. 3. Диаграмма рассеяния линейной регрессионной модели (тип 2)

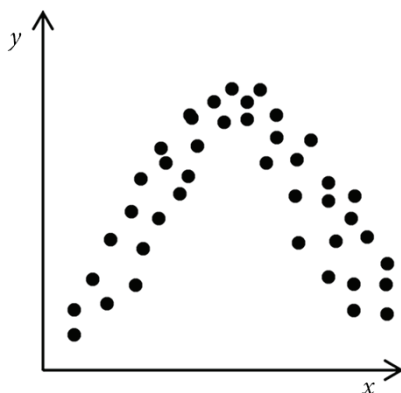


Рис. 4. Диаграмма рассеяния квадратичной регрессионной модели

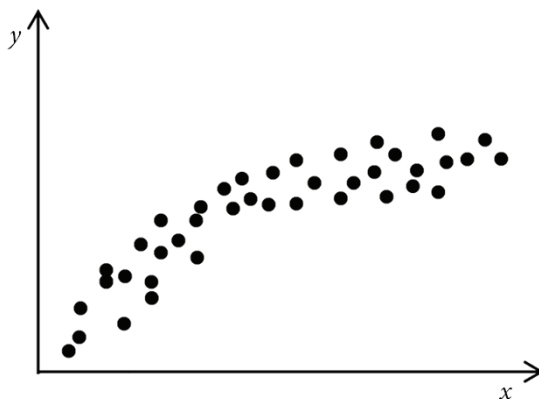


Рис. 5. Диаграмма рассеяния показательной регрессионной модели

4. Степенная функция регрессии (рис. 6):

$$y = \alpha e^{\beta x} \varepsilon.$$

5. Гиперболическая функция регрессии (рис. 7):

$$y = \alpha + \frac{\beta}{x} + \varepsilon.$$

6. x и y могут быть независимыми (рис. 8).

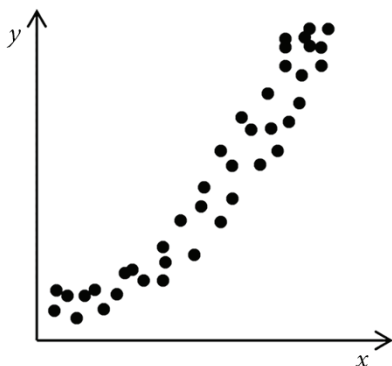


Рис. 6. Диаграмма рассеяния
степенной регрессионной модели

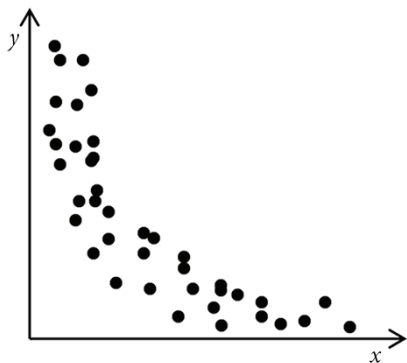


Рис. 7. Диаграмма рассеяния
гиперболической регрессионной
модели

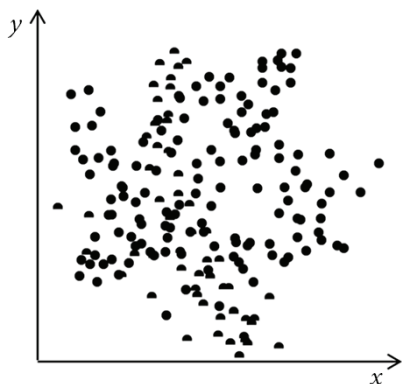


Рис. 8. Диаграмма рассеяния независимых x и y

Эконометрические методы оценки параметров модели

Каждая модель состоит из нескольких уравнений, и каждое уравнение имеет несколько переменных. Сначала рассмотрим модель простой линейной регрессии, которая состоит из двух переменных, чтобы понять технические основы данных нетривиальных методов.

Пусть Y — эндогенная переменная, характеризующая результат или эффективность экономической системы в рамках нашего анализа. Значения Y генерируются внутри системы как следствие других переменных и факторов. Также имеется набор объясняющих (экзогенных) переменных, которые характеризуют состояние экономической системы. Данные переменные объясняют формирование эндогенных переменных.

Рассмотрим рис. 1. Очевидно, что две данные переменные имеют положительную зависимость. Это означает, что увеличение x обычно сопровождается увеличением y . И связь между x и y может быть описана подогнанной прямой линией, соответствующей данным. Конечно, на практике такой метод будет неточным.

Рассмотрим *простую парную модель*:

$$y = \alpha + \beta x + \varepsilon.$$

Переменная y рассматривается как зависимая переменная и состоит из двух составляющих:

1) неслучайной (определенной) составляющей $\alpha + \beta x$, где x — объясняющая (независимая) переменная. Можно отметить, что $y = \alpha + \beta x$ является общим уравнением для прямой линии, которая наилучшим образом соответствует данным;

2) случайной составляющей ε (случайный возмущающий член, ошибка).

Зачем включать случайную составляющую в модель? Существуют следующие важные причины:

1. *Пропущенные переменные*. Есть множество других переменных, которые могут повлиять на y , но они пропущены из-за:

а) *проблемы измерения*. Интуитивно мы знаем, что некоторая переменная влияет на y , но мы не можем включить ее, потому что мы не знаем, как ее измерить (например, психологический фактор или ураган);

б) *низкой значимости фактора*. Мы знаем, как измерить некоторые другие факторы, но их влияние настолько мало, что неразумно включать их в модель;

в) *ошибки исследователя*. Иногда переменная пропускается, потому что исследователь считает, что она незначительная в соответ-

ствии с отсутствием опыта и знаний. Проблема в том, что никогда не известно, какие переменные должны быть включены в модель. Так, мы можем заключить, что вместо зависимости $y = f(X_1, \dots, X_n)$, где n слишком большое для практических целей, мы рассматриваем зависимость с меньшим набором существенных переменных, которые интересуют нас больше всего.

2. *Неправильная функциональная спецификация.* Эта проблема возникает, когда исследователь тестирует модель одного типа, но спецификация более сложна, например, существуют некоторые нелинейные части.

3. *Ошибки в наблюдениях* (например, занижение реальных доходов). В этом случае наблюдаемые данные не отражают реального баланса, а существующие расхождения являются частью случайной составляющей.

Каждое изменение в Y порождается изменением в X и ε .

Продолжим анализ парной линейной регрессионной модели

$$y = \alpha + \beta x + \varepsilon.$$

Графическая иллюстрация подогнанной линии представлена на рис. 9.

Главный вопрос: как правильно оценить α и β ? Интуиция довольно ясна: эти коэффициенты выбираются таким образом, что

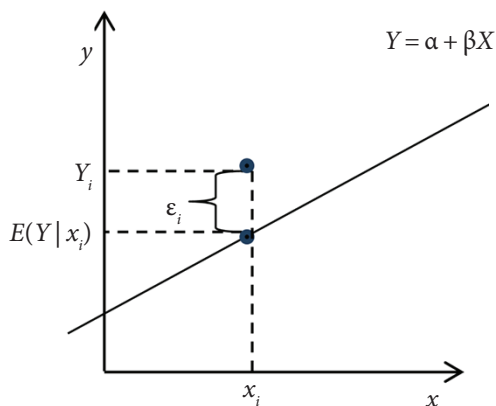


Рис. 9. Графическая иллюстрация подогнанной линии

(по вертикали) расстояния от точек данных до получившейся линии сведены к минимуму (для того чтобы линия максимально приближалась к данным). Это лежит в основе идеи *обычного метода наименьших квадратов* (МНК, OLS).

Этот метод позволяет оценить неизвестные параметры в модели линейной регрессии. Идея метода заключается в минимизации различий между наблюдаемыми результатами в некотором случайном наборе данных и ответами, предсказанными линейной аппроксимацией данных (визуально это рассматривается как сумма вертикальных расстояний между каждой точкой данных в наборе данных и соответствующей точкой на линии регрессии). Чем меньше различия, тем лучше модель описывает данные:

$$Y_i = \alpha + \beta x_i + \varepsilon_i,$$

где ε_i — ошибка.

$\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$ — установленное значение для наблюдения i , где $\hat{\varepsilon}_i$ — остатки для наблюдения i .

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i.$$

Выберем $\hat{\alpha}$ и $\hat{\beta}$ таким образом, чтобы минимизировать

$$\sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2.$$

Графическая иллюстрация МНК дана на рис. 10.

Идея метода МНК: оценить параметры α и β из выборки данных по Y_i и x_i :

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1 \dots N.$$

Минимизируем сумму квадратов остатков:

$$\min_{\hat{\alpha}, \hat{\beta}} S = \sum_i \varepsilon_i^2 = \sum_i (Y_i - \alpha - \beta x_i)^2.$$

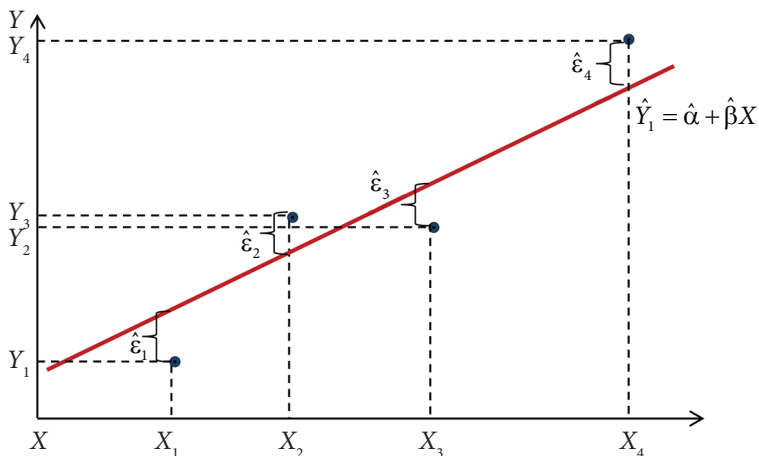


Рис. 10. Графическая иллюстрация обычного метода наименьших квадратов

Условие первого порядка (F.O.C.):

$$\begin{cases} \frac{\partial S}{\partial \hat{\alpha}} = -2 \sum_i (Y_i - \alpha - \beta x_i) = 0 \\ \frac{\partial S}{\partial \hat{\beta}} = -2 \sum_i x_i (Y_i - \alpha - \beta) = 0 \end{cases}$$

После преобразований:

$$\hat{\alpha} = \bar{Y} - \frac{\sum_i (x_i - \bar{x}) Y_i}{\sum_i (x_i - \bar{x})^2} \bar{x},$$

$$\hat{\beta} = \frac{\sum_i (x_i - \bar{x}) Y_i}{\sum_i (x_i - \bar{x})^2}.$$

Для проверки аппроксимации модели используем:

- 1) парный коэффициент корреляции;
- 2) коэффициент детерминации R^2 :

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS},$$

$$TSS = RSS + ESS,$$

где TSS — общая сумма квадратов.

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2.$$

TSS отражает общее изменение переменной y относительно среднего значения.

RSS — сумма квадратов регрессии:

$$RSS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2.$$

RSS отражает объясненную часть вариации переменной y .

ESS — сумма квадратов ошибок:

$$ESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

ESS отражает необъясненную часть вариации переменной y .

R^2 измеряет, насколько хорошо независимые переменные объясняют зависимую переменную Y .

Свойства коэффициента детерминации:

1. $0 \leq R^2 \leq 1$.

Другими словами:

$0 \leq R^2 \leq TSS$.

2. Рассмотрим случай, когда $R^2 = 1$, иначе говоря, $ESS = 0$, следовательно:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = 0,$$

где $i = 1, 2, \dots, N$, $\hat{y}_i = y_i$.

Это означает, что мы нашли модель, которая полностью объясняет наш набор данных, что может произойти только в случае

двух наблюдений. Так, $R^2 = 1$ сигнализирует, что нам необходимо увеличить число наблюдений.

3. $R^2 = 0$, другими словами, $RSS = 0$, следовательно:

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 0,$$

где $i = 1, 2, \dots, N$, $\hat{y}_i = \bar{y}$.

Это означает, что построенная модель не объясняет, что может произойти в следующих случаях:

- а) x и y независимы (рис. 11);
- б) ошибки в спецификации функциональной формы (выбрали линейную зависимость вместо квадратичной; рис. 12);
- в) нет никакой корреляции между переменными, но может быть и другая статистическая связь (например, гетероскедастичность; рис. 13):

4. Модель имеет более высокую объясняющую силу (и \hat{y}_i более точно соответствует y_i), когда R^2 ближе к 1. Это называется хорошим «объяснением», качеством подгонки.

П р и м е ч а н и е. Для парной линейной регрессии $R^2 = r_{xy}^2$.

5. R^2 может увеличиться только в случае добавления новой объясняющей переменной в правую часть уравнения.

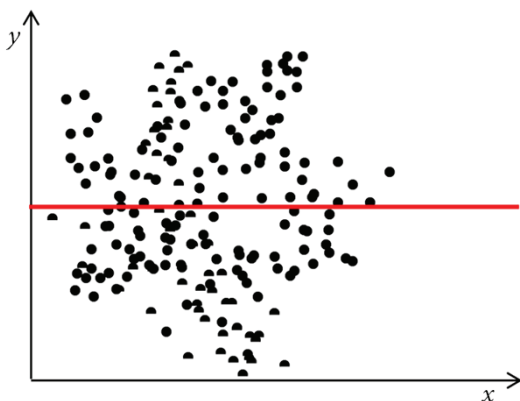


Рис. 11. Графическая иллюстрация независимых переменных

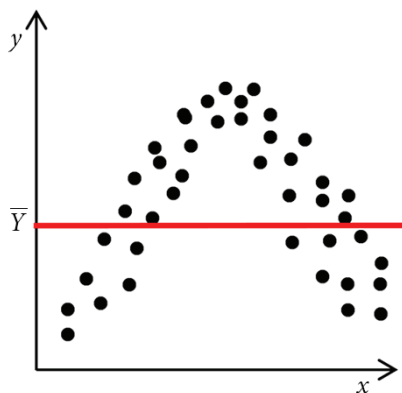


Рис. 12. Графическая иллюстрация ошибок в спецификации функциональной формы

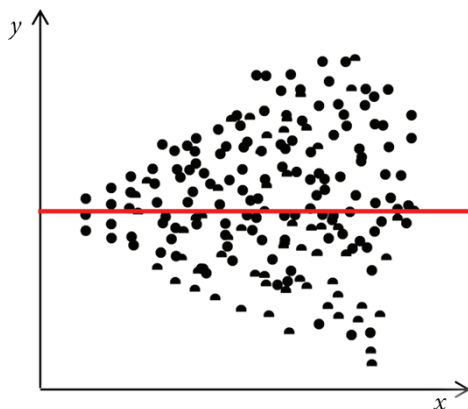


Рис. 13. Графическая иллюстрация гетероскедастичности

Множественная линейная регрессия — это расширение парной линейной регрессии в случае, когда зависимая переменная объясняется более чем одной независимой переменной:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon.$$

Оценка параметров модели реализуется так же, как в модели с парной линейной регрессией (по МНК):

$$\min_{\hat{\beta}_i} \sum_i \varepsilon_i^2,$$

$$\hat{\varepsilon}_i = y_i - \hat{y}_i,$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_k x_{ki},$$

$$\hat{\varepsilon}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki}.$$

Рассмотрим следующее уравнение:

$$\begin{aligned} \min_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} S &= \sum_{i=1}^N \varepsilon_i^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \\ &= \sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki} \right)^2. \end{aligned}$$

Условие первого порядка (F.O.C.):

$$\left\{ \begin{array}{l} \frac{\partial S}{\partial \hat{\beta}_0} = 0 \\ \frac{\partial S}{\partial \hat{\beta}_1} = 0 \\ \dots \\ \frac{\partial S}{\partial \hat{\beta}_k} = 0 \end{array} \right.$$

$$\left\{ \begin{array}{l} \frac{\partial S}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki} \right) = 0(1) \\ \frac{\partial S}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki} \right) x_{1i} = 0(2) \\ \frac{\partial S}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki} \right) x_{2i} = 0(3) \\ \dots \\ \frac{\partial S}{\partial \hat{\beta}_k} = -2 \sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki} \right) x_{ki} = 0(k+1) \end{array} \right.$$

Сделаем преобразование первого уравнения в системе:

$$\sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki} \right) = 0$$

$$\sum_{i=1}^N \left(y_i - \hat{\beta}_0 N - \hat{\beta}_1 \sum_{i=1}^N x_{1i} - \hat{\beta}_2 \sum_{i=1}^N x_{2i} - \dots - \hat{\beta}_k \sum_{i=1}^N x_{ki} \right) = 0$$

$$\hat{\beta}_0 N + \hat{\beta}_1 \sum_{i=1}^N x_{1i} + \hat{\beta}_2 \sum_{i=1}^N x_{2i} + \dots + \hat{\beta}_k \sum_{i=1}^N x_{ki} = \sum_{i=1}^N y_i$$

Сделаем преобразование второго уравнения в системе:

$$\sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki} \right) x_{1i} = 0$$

$$\sum_{i=1}^N y_i x_{1i} - \hat{\beta}_0 \sum_{i=1}^N x_{1i} - \hat{\beta}_1 \sum_{i=1}^N x_{1i}^2 - \hat{\beta}_2 \sum_{i=1}^N x_{1i} x_{2i} - \dots - \hat{\beta}_k \sum_{i=1}^N x_{1i} x_{ki} = 0$$

$$\hat{\beta}_0 \sum_{i=1}^N x_{1i} + \hat{\beta}_1 \sum_{i=1}^N x_{1i}^2 + \hat{\beta}_2 \sum_{i=1}^N x_{1i} x_{2i} + \dots + \hat{\beta}_k \sum_{i=1}^N x_{1i} x_{ki} = \sum_{i=1}^N x_{1i} y_i$$

Сделаем преобразование третьего уравнения в системе:

$$\sum_{i=1}^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki}) x_{2i} = 0$$

$$\sum_{i=1}^N y_i x_{2i} - \hat{\beta}_0 \sum_{i=1}^N x_{2i} - \hat{\beta}_1 \sum_{i=1}^N x_{1i} x_{2i} - \hat{\beta}_2 \sum_{i=1}^N x_{2i}^2 - \dots - \hat{\beta}_k \sum_{i=1}^N x_{2i} x_{ki} = 0$$

$$\hat{\beta}_0 \sum_{i=1}^N x_{2i} + \hat{\beta}_1 \sum_{i=1}^N x_{1i} x_{2i} + \hat{\beta}_2 \sum_{i=1}^N x_{2i}^2 + \dots + \hat{\beta}_k \sum_{i=1}^N x_{2i} x_{ki} = \sum_{i=1}^N x_{2i} y_i$$

...

В результате мы получим следующую систему уравнений:

$$\left\{ \begin{array}{l} \hat{\beta}_0 N + \hat{\beta}_1 \sum_{i=1}^N x_{1i} + \hat{\beta}_2 \sum_{i=1}^N x_{2i} + \dots + \hat{\beta}_k \sum_{i=1}^N x_{ki} = \sum_{i=1}^N y_i \\ \hat{\beta}_0 \sum_{i=1}^N x_{1i} + \hat{\beta}_1 \sum_{i=1}^N x_{1i}^2 + \hat{\beta}_2 \sum_{i=1}^N x_{1i} x_{2i} + \dots + \hat{\beta}_k \sum_{i=1}^N x_{1i} x_{ki} = \sum_{i=1}^N x_{1i} y_i \\ \hat{\beta}_0 \sum_{i=1}^N x_{2i} + \hat{\beta}_1 \sum_{i=1}^N x_{1i} x_{2i} + \hat{\beta}_2 \sum_{i=1}^N x_{2i}^2 + \dots + \hat{\beta}_k \sum_{i=1}^N x_{2i} x_{ki} = \sum_{i=1}^N x_{2i} y_i \\ \dots \\ \hat{\beta}_0 \sum_{i=1}^N x_{ki} + \hat{\beta}_1 \sum_{i=1}^N x_{1i} x_{ki} + \hat{\beta}_2 \sum_{i=1}^N x_{2i} x_{ki} + \dots + \hat{\beta}_k \sum_{i=1}^N x_{ki}^2 = \sum_{i=1}^N x_{ki} y_i \end{array} \right.$$

Эта система называется системой нормальных уравнений для нахождения коэффициентов множественной линейной регрессии с использованием обычного метода наименьших квадратов.

В матричном виде:

$$A \cdot \hat{\beta} = b,$$

$$\hat{\beta} = A^{-1} \cdot b.$$

$$A_{(k+1) \times (k+1)} = \begin{pmatrix} N & \sum_{i=1}^N x_{1i} & \sum_{i=1}^N x_{2i} & \sum_{i=1}^N x_{3i} & \cdots & \sum_{i=1}^N x_{ki} \\ \sum_{i=1}^N x_{1i} & \sum_{i=1}^N x_{1i}^2 & \sum_{i=1}^N x_{1i}x_{2i} & \sum_{i=1}^N x_{1i}x_{3i} & \cdots & \sum_{i=1}^N x_{1i}x_{ki} \\ \sum_{i=1}^N x_{2i} & \sum_{i=1}^N x_{1i}x_{2i} & \sum_{i=1}^N x_{2i}^2 & \sum_{i=1}^N x_{2i}x_{3i} & \cdots & \sum_{i=1}^N x_{2i}x_{ki} \\ \sum_{i=1}^N x_{3i} & \sum_{i=1}^N x_{1i}x_{3i} & \sum_{i=1}^N x_{2i}x_{3i} & \sum_{i=1}^N x_{3i}^2 & \cdots & \sum_{i=1}^N x_{3i}x_{ki} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^N x_{ki} & \sum_{i=1}^N x_{1i}x_{ki} & \sum_{i=1}^N x_{2i}x_{ki} & \sum_{i=1}^N x_{3i}x_{ki} & \cdots & \sum_{i=1}^N x_{ki}^2 \end{pmatrix}$$

$$\hat{\beta}_{(k+1) \times 1} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \cdots \\ \hat{\beta}_k \end{pmatrix}$$

$$b_{(k+1) \times 1} = \begin{pmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_{1i} y_i \\ \sum_{i=1}^N x_{2i} y_i \\ \sum_{i=1}^N x_{3i} y_i \\ \cdots \\ \sum_{i=1}^N x_{ki} y_i \end{pmatrix}$$

Введем обозначения:

X — матрица наблюдений независимых переменных.

Необходимо ввести коэффициент, состоящий из единичных элементов, поскольку в уравнении есть свободный коэффициент:

$$y = \beta_0 \cdot 1 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon.$$

$$X_{N \times (k+1)} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & x_{32} & \dots & x_{k2} \\ 1 & x_{13} & x_{23} & x_{33} & \dots & x_{k3} \\ 1 & x_{14} & x_{24} & x_{34} & \dots & x_{k4} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1N} & x_{2N} & x_{3N} & \dots & x_{kN} \end{pmatrix}$$

Вектор зависимой переменной наблюдений:

$$y_{N \times 1} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_N \end{pmatrix},$$

$$A = (x^T x),$$

$$b = x^T y.$$

Система нормальных уравнений для нахождения коэффициентов множественной линейной регрессии с использованием обычного метода наименьших квадратов в матричном виде выглядит следующим образом:

$$(x^T x) \hat{\beta} = x^T y.$$

Оценка коэффициентов модели после реализации обычного метода наименьших квадратов представлена следующим образом:

$$\hat{\beta}_{OLS} = (x^T x)^{-1} x^T y.$$

Для получения наилучших результатов по методу МНК должны выполняться следующие условия (условия Гаусса — Маркова):

1. *Линейность по параметрам.* Спецификация модели выглядит следующим образом:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon,$$

где $\beta_1, \beta_2, \dots, \beta_k$ — интересующие нас коэффициенты; ε — ошибка.

2. *Случайная выборка.* Выборка объясняющих переменных (x_1, x_2, \dots, x_k) должна быть случайной.

3. *Отсутствие полной коллинеарности.* Нет объясняющих переменных, которые являются константами, и нет линейной зависимости между объясняющими переменными.

4. *Нулевое среднее значение (условное).* Ошибка ε имеет нулевое ожидаемое значение при любых значениях независимых переменных:

$$E(\varepsilon | x_1, x_2, \dots, x_k) = E(\varepsilon) = 0.$$

5. *Гомоскедастичность.* Ошибка ε имеет постоянную дисперсию при любом значении объясняющих переменных (другими словами, дисперсия ошибки не зависит от количества наблюдений):

$$\text{var}(\varepsilon | x_1, x_2, \dots, x_k) = \sigma^2.$$

6. *Отсутствие автокорреляции.* Ошибки для разных наблюдений не коррелируют друг с другом:

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \text{ если } i \neq j.$$

7. *Нормальность.* Ошибки имеют нормальное распределение:

$$\varepsilon_i \sim N(0, \sigma^2).$$

Условия 1–5 известны как *условия Гаусса — Маркова*. Если данные условия выполняются в модели, то такая модель называется классической множественной линейной регрессионной моделью. Также если выполняется условие нормальности, то такая модель

называется нормальной множественной линейной регрессионной моделью.

Теорема Гаусса — Маркова. Если мы ограничим класс конкурирующих оценок должным образом, то можно показать, что МНК является лучшим в этом классе. По предположениям 1–5 МНК-оценка $\hat{\beta}_j$ для β_j — *наилучшая линейная несмещенная оценка (BLUE)*.

Фиктивные переменные используются для качественных данных. Это касается чего-то, что нельзя посчитать просто. Фиктивные переменные могут иметь только два типа значений: «0» и «1». При определении фиктивной переменной мы должны решить, какое событие будет иметь значение 1 и какое равно 0. Например, при анализе индивидуальной заработной платы мы можем определить, что пол является бинарной переменной, принимающей значение 1 для женщин и значение 0 для мужчин. Название переменной в данном случае указывает на событие со значением один. Аналогично информация фиксируется для определения мужчины, когда переменная принимает значение 1, если человек является мужчиной, и 0, если лицо является женщиной.

Если модель имеет признак, который может принимать несколько значений, тогда нам нужно использовать общее правило включения фиктивных переменных: общее число фиктивных переменных должно быть $(k - 1)$, где k — количество возможных значений качественного признака, если в модели есть константа. Необходимо избегать совершенной коллинеарности между переменными.

Пример. Мы хотим проанализировать, в какое время года спрос на мороженое является самым высоким. Для этого нам нужно ввести четыре фиктивные переменные:

$$d_{summer} = \begin{cases} 1, & \text{если время года лето} \\ 0, & \text{если зима, осень или весна} \end{cases}$$

$$d_{spring} = \begin{cases} 1, & \text{если весна} \\ 0, & \text{если зима, осень или лето} \end{cases}$$

$$d_{autumn} = \begin{cases} 1, & \text{если осень} \\ 0, & \text{если зима, весна или лето} \end{cases}$$

$$d_{winter} = \begin{cases} 1, & \text{если зима} \\ 0, & \text{если осень, весна или лето} \end{cases}$$

Тогда мы получаем следующую модель:

$$y = \beta_0 + \beta_1 d_{summer} + \beta_2 d_{autumn} + \beta_3 d_{winter} + \beta_4 d_{spring} + \varepsilon.$$

Набор данных представлен в табл. 1, средние данные приводятся в табл. 2.

Таблица 1

Набор данных для анализа

Месяц	d_{summer}	d_{spring}	d_{autumn}	d_{winter}
Январь	0	0	0	1
Февраль	0	0	0	1
Март	0	1	0	0
Апрель	0	1	0	0
Май	0	1	0	0
Июнь	1	0	0	0
Июль	1	0	0	0
Август	1	0	0	0
Сентябрь	0	0	1	0
Октябрь	0	0	1	0
Ноябрь	0	0	1	0
Декабрь	0	0	0	1

Таблица 2

Средний спрос на мороженое в каждое время года

Время года	Средний спрос на мороженое
Лето	$\beta_0 + \beta_1$
Осень	$\beta_0 + \beta_2$
Зима	$\beta_0 + \beta_3$
Весна	$\beta_0 + \beta_4$

Однако

$$d_{summer} + d_{autumn} + d_{winter} + d_{spring} = 1,$$

поэтому если данные верны, мы не можем оценить модель.

Вот почему нам нужно принять во внимание правило о введении некоторых фиктивных переменных одного и того же признака и исключить любую переменную, например, весну:

$$y = \beta_0 + \beta_1 d_{summer} + \beta_2 d_{autumn} + \beta_3 d_{winter} + \varepsilon.$$

В полученной модели нет проблемы мультиколлинеарности. Поэтому после таких преобразований мы можем оценить модель.

Тестирование статистических гипотез

Существуют два основных инструмента для тестирования статистических гипотез. Они могут использоваться для вычисления тестовой статистики и доверительных интервалов. В обоих тестах мы исходим из данной гипотезы, а именно нулевой. Статистика теста вычисляется исходя из предположения, что справедлива нулевая гипотеза. Решение о том, что расчетное значение тестовой статистики маловероятно относится к известному распределению, указывает на то, что нулевая гипотеза вряд ли выполняется.

T-тест используется для проверки гипотез об одном параметре: тестирование на односторонние и на двухсторонние альтернативы. *F-тест* имеет разные цели и результаты и используется для тестирования нескольких линейных ограничений: проверки ограничений на исключение и *F*-статистики для общей значимости регрессии.

Иногда выявленная проблема требует ответов на вопрос не только о влиянии одной переменной, но и о влиянии связанного набора переменных. Сравните: влияние курения на здоровье (эффект одной переменной) и половозрастное влияние на ожидаемую заработную плату (эффект двух переменных).

T-тест проводится для гипотезы с одним параметром. Важно, что мы не знаем точных параметров $\beta_1, \beta_2, \dots, \beta_k$, но мы можем

сформулировать некоторую гипотезу об одном из них и проверить ее с использованием статистических методов. Предположим, мы имеем модель

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon,$$

удовлетворяющую условиям 1–7.

Исходим из нулевой гипотезы:

$$H_0 : \beta_j = 0.$$

Это означает, что если мы не отвергаем H_0 с учетом всех $x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_k$, то x_j не влияет на ожидаемое значение y .

Пр и м е р (уравнение заработной платы). Предположим, что у нас есть модель, где зависимая переменная $\log(wage)$ — логарифм заработной платы, а факторами являются уровень образования (*educ*), опыт работы (*exper*) и стаж (*tenure*):

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \varepsilon.$$

Нулевая гипотеза $H_0 : \beta_2 = 0$ означает, что если факторы образования и стажа были учтены, то опыт работы человека (переменная *exper*) не влияет на размер заработной платы. Если эта гипотеза справедлива, значит, опыт работы человека до текущей занятости не влияет на заработную плату. Если $\beta_2 > 0$, то опыт работы положительно влияет на заработную плату.

Какую статистику использовать для проверки гипотезы H_0 ? Данная статистика называется *t*-статистика или *t*-коэффициент и определяется как

$$t = \frac{\hat{\beta}_j - \beta_j}{s.e.(\hat{\beta}_j)}.$$

Это отношение стандартной нормальной переменной и квадратного корня из независимой хи-квадратной переменной и, следовательно, соответствует *t*-распределению Стьюдента с $N - K$ степенями свободы. При этом *t*-распределение близко к стандартному нормальному распределению, за исключением того, что оно имеет более плоские хвосты, особенно когда количество степеней свободы $N - K$

мало. Чем больше $N - K$, тем t -распределение больше соответствует стандартному нормальному, а при достаточно большом $N - K$ оба распределения идентичны.

По условиям 1–6 распределение выборки тестовой t -статистики:

$$\frac{\hat{\beta}_j - \beta_j}{s.e.(\hat{\beta}_j)} \sim t_{n-k-1},$$

где n — количество наблюдений; k — количество включенных переменных.

Другим важным аспектом тестирования гипотез является выбор уровня значимости α — вероятности отвержения нулевой гипотезы, когда нулевая гипотеза действительно верна (как правило, $\alpha = 0,01$, $0,05$ или $0,10$).

Существует два типа ошибок:

- *ошибка I типа*: отвержение H_0 , когда она на самом деле верна;
- *ошибка II типа*: неспособность отклонить H_0 , когда она не верна.

Тогда вероятность совершения ошибки I типа определяется как уровень значимости теста и обозначается как α .

Классический пример: нулевая гипотеза предполагает, что человек виновен. Тогда ошибка I типа означает, что лицо будет признано невиновным, а на самом деле является виновным, а ошибка II типа — лицо будет осуждено за преступление при действительной невиновности.

Вероятность ошибки I типа напрямую контролируется исследователем посредством его выбора уровня значимости α . Когда тест выполняется на 5 %-м уровне, вероятность отклонения нулевой гипотезы при ее справедливости равна 5 %. Данная вероятность (уровень значимости) часто упоминается как размер теста. Вероятность ошибки II типа зависит от истинных значений параметров. Если истина сильно отклоняется от указанной нулевой гипотезы, вероятность такой ошибки будет относительно небольшой, в то время как если нулевая гипотеза близка к истине, вероятность ошибки будет довольно большой. Обратная вероятность, т.е. вероятность отвергнуть нулевую гипотезу, когда она неверна, называется мощ-

ностью теста. Она указывает, насколько «мощным» является тест в нахождении отклонений от нулевой гипотезы (в зависимости от истинного значения параметра). В общем случае уменьшение размера теста уменьшает его мощность, поэтому существует компромисс между ошибками I и II типа.

Выделяют два типа t -теста:

1. Двусторонний t -тест:

$$H_0 : \beta = 0,$$

$$H_a : \beta \neq 0.$$

Предположим, что уровень значимости $\alpha = 0,05$. $t_{\hat{\beta}_j}$ должно быть «достаточно большим», чтобы отклонить H_0 в пользу H_a . «Достаточно большое» $t_{\hat{\beta}_j}$ при 5 %-м уровне значимости на 95 % в t -распределении с $n - k - 1$ степенями свободы обозначим как t_{cr} (t критическое). Правило отклонения: H_0 отклоняется в пользу H_a на уровне значимости 5 %, если $\left| t_{\hat{\beta}_j} \right| > t_{cr}$.

Значение t_{cr} определяется с помощью специальной таблицы распределения Стьюдента. Область отклонения отмечена на рис. 14.

2. Односторонний t -тест:

$$H_0 : \beta = 0,$$

$$H_a : \beta < 0 \text{ или } H_a : \beta > 0.$$

Предположим, что уровень значимости $\alpha = 0,05$. $t_{\hat{\beta}_j}$ должно быть «достаточно большим», чтобы отклонить H_0 в пользу H_a . Правило отклонения: H_0 отклоняется в пользу H_a на уровне значимости 5 %, если $t_{\hat{\beta}_j} > t_{cr}$ (для $H : \beta > 0$) и $t_{\hat{\beta}_j} < -t_{cr}$ (для $H_a : \beta < 0$). Область отклонения отмечена на рис. 15.

На практике для проверки гипотез обычно смотрят p -значение, имеющее наименьший уровень значимости, на котором отклоняется нулевая гипотеза.

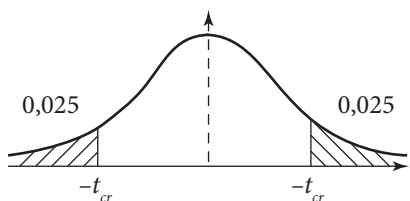


Рис. 14. Графическая иллюстрация двухстороннего t -теста с уровнем значимости 5 %

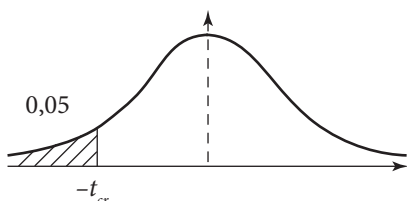


Рис. 15. Графическая иллюстрация одностороннего t -теста ($H_a : \beta < 0$) с уровнем значимости 5 %

F -тест используется для проверки того, имеет или не имеет группа переменных влияния на зависимую переменную. Рассмотрим следующий пример:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \\ + \beta_3 gender + \beta_4 marital + \beta_5 children + \varepsilon,$$

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0,$$

$$H_a : \text{по крайней мере из } \beta_3, \beta_4, \beta_5 \neq 0.$$

Тест, который мы рассматриваем здесь, построен для обнаружения любого нарушения H_0 . Также тест выполняется, когда альтернатива аналогична $H_a : \beta_3 > 0$, или $\beta_4 > 0$, или $\beta_5 > 0$, хотя он не будет наилучшим тестом при таких альтернативах. У нас нет пространственных или статистических предпосылок, необходимых для охвата тестов, которые обладают большей мощностью по нескольким односторонним альтернативам. Нулевая гипотеза имеет три ограничения на исключение: если мы не отвергаем ее, то пол (*gender*), семейное положение (*marital*) и количество детей (*children*) не влияют на заработную плату (*wage*), а образование (*educ*) и опыт *exper* учитываются и поэтому должны быть исключены из модели. Проверка такой гипотезы называется совместным тестом на гипотезу. Модель с тремя данными переменными называется неограниченной моделью, а без трех данных переменных — ограниченной. Ограниченная модель выглядит следующим образом:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \varepsilon.$$

Статистика теста для данного случая называется F -статистикой и определяется как

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n - k - 1)},$$

где SSR_R — сумма квадратов остатков из ограниченной модели;

SSR_{UR} — сумма квадратов остатков из неограниченной модели;

q — количество ограничений;

$n - k - 1$ — степень свободы в неограниченной модели.

Итак, сначала мы получаем степени свободы в неограниченной модели, затем решаем, сколько переменных должно быть исключено в ограниченной модели; это q . SSR сообщаются с каждой МНК-регрессией, поэтому формирование F -статистики является простым. Чтобы использовать F -статистику, мы должны знать ее выборочное распределение при нулевой гипотезе для того, чтобы выбрать критические значения и правила отклонения. Можно показать, что при H_0 (и предполагая, что указанные условия выполнены) F распределяется как случайная величина F с $(q, n - k - 1)$ степенями свободы. Можно представить это как

$$F \sim F_{q, n-k-1}.$$

Правило отклонения: $F > F_{cr}$.

Знание суммы квадратов остатков ничего не говорит нам о справедливости гипотезы H_0 . Но дело в том, что нам важно, насколько возрастает SSR , когда мы исключаем рассматриваемые переменные из модели. Помните, что МНК-оценки выбираются так, чтобы минимизировать сумму квадратов остатков, SSR всегда увеличивается, когда переменные отбрасываются из модели; это алгебраический факт. Вопрос заключается в том, достаточно ли это увеличение относительно SSR в модели со всеми переменными, чтобы оправдать отклонение нулевой гипотезы.

Тест нескольких ограничений называется *множественной проверкой гипотез* или *совместным тестом на гипотезы*. Если H_0 отклоняется, то это означает, что переменные в нулевой гипотезе являются совместно значимыми.

Другой случай F -теста — *тестирование на общую значимость регрессии*. Данный тест может быть представлен следующим образом:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \text{ (неограниченная модель),}$$

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_k = 0,$$

$$H_a : \text{по крайней мере одна из них} \neq 0,$$

$$y = \beta_0 + \varepsilon \text{ (ограниченная модель).}$$

Соответствующая тестовая статистика в этом случае:

$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}.$$

Правило отклонения: $F > F_{cr}$.

Конечная вероятность, играющая роль в статистических тестах, обычно называется p -значением. Данная p или значение вероятности обозначает минимальный размер, для которого нулевая гипотеза будет отвергнута. Она определяется как вероятность, при справедливости нулевой гипотезы, превышения тестовой статистики (в абсолютном значении) над значением статистики, вычисленной по выборке. Если p -значение меньше уровня значимости α , нулевая гипотеза отклоняется. Многие современные программные пакеты предоставляют такие p -значения и позволяют исследователям делать свои выводы без консультаций или вычисления соответствующих критических значений. Это значение также показывает чувствительность решения об отклонении нулевой гипотезы с учетом уровня значимости. Например, p -значение 0,08 указывает, что нулевая гипотеза отклоняется на уровне значимости 10 %, но не на уровне 5 %.

В заключение можно сказать, что тестирование гипотез используется для ответа на конкретные вопросы о влиянии объясняющих переменных на зависимую переменную. Также t -тест используется для проверки гипотез относительно конкретного коэффициента, а F -тест — для проверки того, имеет или не имеет группа переменных влияние на зависимую переменную.

ГЛАВА 2. НАРУШЕНИЕ ОСНОВНЫХ ПРЕДПОЛОЖЕНИЙ МНК-ОЦЕНКИ

Гетероскедастичность

Если некоторые из указанных условий Гаусса — Маркова (1–6) нарушены, то возникают различные проблемы, которые ведут к смещенным или несостоятельным оценкам, значит, результаты оценок будут неверны. Основными проблемами, обуславливающими такие нарушения, являются: *гетероскедастичность, мультиколлинеарность, автокорреляция и эндогенность*. Исследователи прикладывают много усилий для решения этих проблем и получения разумных и правильных результатов оценки. Рассмотрим данные нарушения более подробно.

Пятое условие Гаусса — Маркова представляет собой гомоскедастичность, которая утверждает, что ошибки ε имеют постоянную дисперсию при любом значении объясняющей переменной. Это означает, что дисперсия ошибок постоянна во времени для всех наблюдений:

$$\text{var}(\varepsilon_i | x_1, x_2, \dots, x_k) = \sigma^2 = \text{const.}$$

Нарушение этого предположения ведет к проблеме гетероскедастичности, когда дисперсия меняется все время:

$$\text{var}(\varepsilon_i | x_1, x_2, \dots, x_k) = \sigma^2 \neq \text{const.}$$

Это тот случай, когда вероятность того, что ε_i в выборке принимает определенные положительные (отрицательные) значения, различна для разных наблюдений (рис. 16, 17).

Существуют следующие *причины гетероскедастичности*:

1. Переменные имеют различные величины для различных наблюдений.

Пример. Оценивается влияние увеличения доходов на образовательные расходы населения. В денежном выражении 1 % изменение дохода жителей Москвы (большой город) полностью отлича-

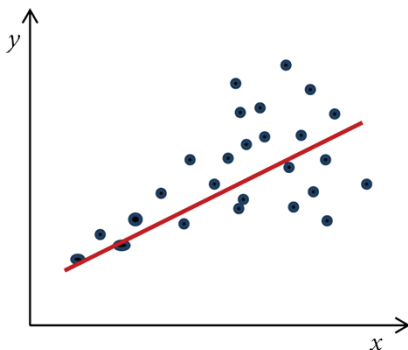


Рис. 16. Графическая иллюстрация гетероскедастичности

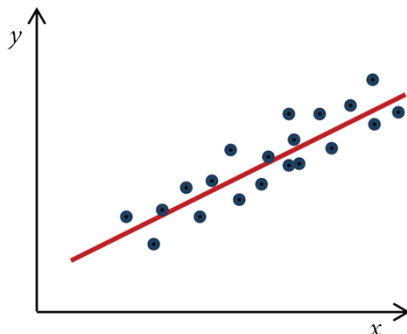


Рис. 17. Графическая иллюстрация гомоскедастичности

ется от того же 1 % изменения дохода населения самой маленькой сельской местности.

2. В случае временных рядов x и y увеличиваются с течением времени. Это влияет на дисперсию ошибки.

3. Неправильная спецификация модели.

Пример: рассматриваем линейную модель, но на самом деле она логарифмическая.

Гетероскедастичность имеет место, когда дисперсия ошибок не постоянна по наблюдениям. Это делает стандартные ошибки неправильными.

Существует три основных примера гетероскедастичности ошибок:

1) σ_e^2 увеличивается одновременно с увеличением переменной X , (представлен на рис. 18, а);

2) σ_e^2 имеет самые высокие значения, когда X принимает средние значения, и снижается при приближении к экстремальным значениям переменной (рис. 18, б);

3) σ_e^2 имеет наивысшие значения, когда значения X являются самыми низкими, и быстро уменьшается и становится однородным при увеличении X (рис. 18, в).

Истинная гетероскедастичность возникает из-за изменчивости случайного члена и из-за ее зависимости от разных факторов. Ложная гетероскедастичность очевидна при наличии ошибок специфици-

кации (например, если истинная модель является логарифмической, а мы оцениваем линейную).

Последствия гетероскедастичности:

1) оценка остается несмещенной, но перестает быть эффективной;

2) не можем применять t -тест и F -тест.

Если у нас есть регрессия $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$, проблема гетероскедастичности возникает при $E(\varepsilon^2 | x_1, x_2, \dots, x_k) = \sigma_x^2$, а при гомоскедастичности $\text{var}(\varepsilon | x_1, x_2, \dots, x_k) = \sigma^2$ или $E(\varepsilon^2 | x_1, x_2, \dots, x_k) = E(\varepsilon^2) = \sigma^2$.

Гетероскедастичность можно обнаружить, если использовать частный и общий тесты. Частный тест — тест Голдфелда — Квандта, основанный на нормальном распределении случайной ошибки в моделях регрессии. В условие этого теста включаем $H_0: \text{var}(\varepsilon | x_1, x_2, \dots, x_k) = \sigma^2$, H_a : гетероскедастичность в модели. Для начала данные упорядочиваются в порядке убывания независимой переменной Z . Тогда осуществляется обычная МНК-оценка исходной модели регрессии для двух различных выборок, первого и последнего m наблюдений в этом порядке, где $m < n/2$. Для полученных двух оценок регрессионной модели вычисляется F -статистика:

$$F = \frac{ESS_1 / (m - k)}{ESS_2 / (m - k)},$$

где m — количество наблюдений; k — количество переменных; ESS — сумма квадратов отклонений. Данная статистика при отсутствии

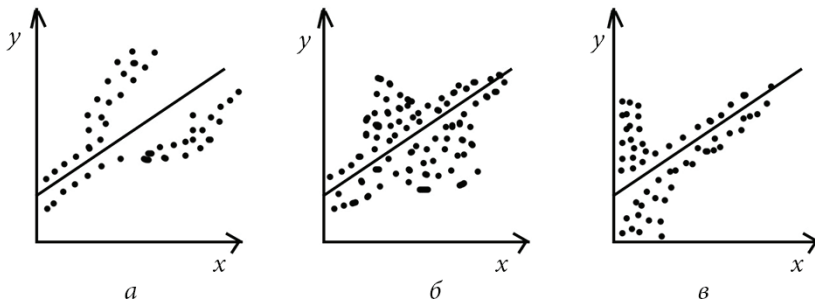


Рис. 18. Примеры гетероскедастичности ошибок

гетероскедастичности и нормальности распределения ошибок имеет распределение Фишера: $F(m - k, m - k)$.

Если $F > F_{cr}$, то отвергаем H_0 , и это означает, что в модели существует гетероскедастичность. Если $F < F_{cr}$, то гетероскедастичность данного типа признается несущественной.

Другим способом тестирования на гетероскедастичность является тест Бройша — Пагана. Предположим, что имеется регрессия специальной формы ε :

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon,$$

$$\varepsilon^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k + v.$$

H_0 : гомоскедастичность, т.е. $\delta_1 = \delta_2 = \dots = \delta_k = 0$.

H_a : по крайней мере одна из $\delta_i \neq 0$.

Начинаем с оценки линейной регрессии по МНК и получаем $\hat{\varepsilon}^2$, затем, используя $\hat{\varepsilon}^2$, оцениваем уравнение для ε^2 . Далее вычисляем R^2 , а потом сравниваем с тестовой статистикой $nR^2 \sim \chi_k^2$, где n — количество наблюдений и χ_k^2 — распределение Стьюдента. Если $nR^2 > \chi_{k(cr)}^2$, то отвергаем H_0 .

Иногда оба типа тестов дают разные ответы. Частные тесты проверяют только особый случай гетероскедастичности, но они более мощные. Общие анализы имеют широкий диапазон проверок, но они менее мощные. Если частный тест обнаруживает проблему, а общий — нет, то в модели существует гетероскедастичность. Если общий тест обнаруживает гетероскедастичность, а частный — нет, мы не можем сказать чего-то определенного. В этом случае мы должны использовать тест Уайта. Для этого мы генерируем вспомогательную регрессию:

$$\begin{aligned} \hat{\varepsilon}^2 = & \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_1^2 + \\ & + \delta_5 x_2^2 + \delta_6 x_3^2 + \delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + v. \end{aligned}$$

Предположение о гомоскедастичности, $\text{var}(\varepsilon | x_1, x_2, \dots, x_k) = \sigma^2$, может быть заменено предположением, что ε^2 не коррелирует со всеми независимыми переменными (x_j), квадратами независимых

переменных (x_j^2) , всеми перекрестными продуктами $(x_j, x_h, j \neq h)$. Формальная процедура теста Уайта следующая:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon.$$

Оценим уравнение регрессии по МНК. Получим \hat{y} и $\hat{\varepsilon}$. Вычислим \hat{y}^2 и $\hat{\varepsilon}^2$.

Запишем регрессию:

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + v.$$

Получим R^2 и затем сравним с тестовой статистикой $nR^2 \sim \chi_k^2$: если $nR^2 > \chi_{k(cr)}^2$, то мы отклоняем H_0 .

После выявления гетероскедастичности в данных мы можем исправить ее, изменяя масштаб переменных. Если этот метод не работает, мы можем применить робастные (надежные) ошибки. Использование таких робастных стандартных ошибок стало обычной практикой во многих областях применения. Поскольку полученная тестовая статистика (асимптотически) подходящая, независимо от того, имеют ли ошибки постоянную дисперсию, это называется «гетероскедастически надежным умозаключением». Если у вас есть представление о форме гетероскедастичности, осуществление обобщенного метода наименьших квадратов может обеспечить более эффективную оценку. В следующих двух подразделах приводятся соответствующие примеры.

В табл. 3 приводятся отношения правдоподобия и тестовой статистики Вальда, которые позволяют проверить гипотезу о том, что все коэффициенты в модели, кроме точки пересечения, равны нулю. Тест Вальда базируется на робастной матрице ковариаций и, следовательно, является более целесообразным, чем тест отношения правдоподобия, который предполагает, что распределение Пуассона является правильным. Тест Вальда решительно отвергает гипотезу о том, что условное среднее постоянно и не зависит от объясняющих переменных. Псевдо- R^2 представлен в таблице как индекс отношения правдоподобия, поскольку он вычисляется многими пакетами программного обеспечения. Как и во всех нелинейных моделях, здесь не существует универсального определения критерия, когда счетные данные хорошо объясняются моделью.

Таблица 3

Результаты оценки модели Пуассона, *MLE* и *QMLE* [1, с. 69]

Показатель	<i>MLE</i>		<i>QMLE</i>
	оценка	стандартная ошибка	устойчивая стандартная ошибка
const	−0,8737	0,0659	0,7429
log(R&D)	0,8545	0,0084	0,0937
<i>aerospace</i>	−1,4218	0,0956	0,3802
<i>chemistry</i>	0,6363	0,0255	0,2254
<i>computers</i>	0,5953	0,0233	0,3008
<i>machines</i>	0,6890	0,0383	0,4147
<i>vehicles</i>	−1,5297	0,0419	0,2807
<i>Japan</i>	0,2222	0,0275	0,3528
<i>USA</i>	−0,2995	0,0253	0,2736
<i>loglikelihood</i>	−4950,789		
<i>Pseudo R2</i>	0,675		
<i>LR test</i>	20587.54 ($p = 0,000$)		
<i>Wald test</i>			338,9 ($p = 0,000$)

Последствия гетероскедастичности:

1. Коэффициенты, оцениваемые по методу наименьших квадратов, являются несмещенными.
2. Существуют неправильные стандартные ошибки оценок. Следовательно, мы не можем использовать метод наименьших квадратов при выполнении *t*- и *F*-тестов и построения доверительных интервалов.

Как определить проблему гетероскедастичности? Существуют некоторые тесты для обнаружения гетероскедастичности. Наиболее распространенными являются тесты Бройша — Пагана и Уайта. Также мы можем использовать тест Спирмена, тест Голдфелда — Квандта и Глейзера. Более подробно тестирование гетероскедастичности описано далее.

Если мы выявили проблему гетероскедастичности в модели, то необходимо сделать коррекцию. Для этого мы можем использовать робастные стандартные ошибки или масштабные переменные (например, переменные на душу населения).

Автокорреляция

Шестое условие Гаусса — Маркова предполагает отсутствие автокорреляции:

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \text{ если } i \neq j.$$

В результате нарушения данного предположения возникает проблема автокорреляции:

$$\text{cov}(\varepsilon_i, \varepsilon_j) \neq 0, \text{ если } i \neq j.$$

Это обычный случай для оценки временных рядов.

Причины автокорреляции:

1. Пропущенные релевантные переменные. В результате они могут коррелировать с ошибками (лаги включенных переменных, тренды, сезонные эффекты).

2. Высокая частота данных.

3. Неправильная спецификация модели.

Последствия автокорреляции:

1. Оценки по МНК являются смещенными и неэффективными: оценки β_i имеют дисперсию, которая не является самой низкой. Следовательно, эти оценки не являются наиболее точными.

2. Неправильные стандартные ошибки оценок. Следовательно, мы не можем использовать их при выполнении t - и F -тестов и построения доверительных интервалов.

Существует несколько тестов для определения автокорреляции:

а) t -тест с $AR(1)$.

Предположим, существует модель

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t,$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + e_t.$$

Прежде всего нужно выполнить МНК y_t на x_{t1}, \dots, x_{tk} . Затем мы должны получить остатки $\hat{\varepsilon}_t$ и после этого выполнить МНК $\hat{\varepsilon}_t$ на $\hat{\varepsilon}_{t-1}$. И последний этап — выполнить следующий t -тест:

$$H_0 : \rho = 0,$$

$$H_a : \rho \neq 0.$$

В случае отклонения H_0 можем сделать вывод, что в модели имеется последовательная корреляция (автокорреляция);

б) тест Дарбина — Уотсона — это тестовая статистика, которая обнаруживает наличие автокорреляции в остатках.

Предположим, что у нас есть модель

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t.$$

Прежде всего мы должны выполнить МНК y_t на x_{t1}, \dots, x_{tk} , тогда получим остатки $\hat{\varepsilon}_t$. Вторым шагом является вычисление статистики Дарбина — Уотсона (DW):

$$DW = \frac{\sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2},$$

$$DW \approx 2(1 - \hat{\rho}).$$

После этого нужно сформулировать H_0 и H_a :

$$H_0 : \rho = 0 \text{ (нет автокорреляции),}$$

$$H_a : \rho > 0.$$

DW сравнивается с двумя наборами критических значений: d_u и d_L .

Правило отклонения для этого случая следующее:

- если $DW < d_p$, то H_0 отвергается,
- если $DW > d_u$, то нет оснований отвергнуть H_0 ,
- если $d_l < DW < d_u$, то тест не дает ответа;

в) тест Бройша — Годфри является тестом для выявления автокорреляции более высокого порядка, он более общий, чем тест Дарбина — Уотсона, который может тестировать только авторегрессионную модель первого порядка.

Предположим, что мы имеем модель

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t,$$

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_q \varepsilon_{t-q} + e_t.$$

Первым шагом в тесте Бройша — Годфри выполняется МНК y_t на x_{t1}, \dots, x_{tk} , затем получаем остатки $\hat{\varepsilon}_t$.

Вторым шагом выполняется МНК $\hat{\varepsilon}_t$ на $x_{t1}, \dots, x_{tk}, \hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}, \dots, \hat{\varepsilon}_{t-q}$.

Следующий этап — осуществление F -теста:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_q = 0,$$

$$H_a : \text{хотя бы один из них} \neq 0.$$

После этого нужно вычислить LM -статистику:

$$LM = (n - q) R_{\varepsilon}^2.$$

Правило отклонения следующее: если $LM > \chi_q^2$, то H_0 отклоняется, это означает, что существует автокорреляция в модели.

Коррекция автокорреляции. Если мы выявили автокорреляцию в модели, то необходимо сделать коррекцию. Для этого мы можем использовать следующие способы:

- 1) включить пропущенные переменные (лаги, тренд, контроль сезонности);
- 2) применить более подходящие методы для данных с высокой частотой;
- 3) изменить спецификацию модели.

Мультиколлинеарность

Еще одно нарушение условий Гаусса — Маркова — *мультиколлинеарность*. Это ситуация, когда существует высокая корреляция между независимыми переменными. Часто это проблема выборки, а не общего множества. Когда в модели существует мультиколлинеарность, мы можем оценить только общее влияние переменных на y . Мы не можем оценить влияние каждой переменной по отдельности.

Не существует тестов для выявления мультиколлинеарности. Мультиколлинеарность возникает из-за высокой корреляции между объясняющими переменными. Исследователь может попытаться найти нарушение условия Гаусса — Маркова, учитывая следующие моменты:

1. При добавлении или удалении прогностической переменной происходят значительные изменения в оценочных коэффициентах регрессии.
2. Существуют незначимые коэффициенты регрессии для объясняющих переменных в регрессии, в то время как отвержение совместной гипотезы показало, что эти коэффициенты равны нулю (выполнение F -теста).
3. Если имеется незначимый коэффициент для конкретной объясняющей переменной, но простая линейная регрессия объясненной переменной по данной объясняющей переменной показывает, что ее коэффициент существенно отличается от нуля, эта ситуация указывает на мультиколлинеарность во множественной регрессии. Строим матрицу парных коэффициентов корреляции. Случай, когда $|r_{x_1x_2}| > 0$, можно интерпретировать как проблему мультиколлинеарности.

Последствия мультиколлинеарности:

1. Оценочные коэффициенты значительно меняются, когда спецификации модели или данные меняются незначительно. Изменения в данных приводят к изменениям знаков.
2. В случае совершенной коллинеарности существует неспособность оценить коэффициенты для некоторых переменных.

П р и м е ч а н и я: 1) совершенная коллинеарность: есть точные взаимосвязи между объясняющими переменными; 2) несовершен-

ная коллинеарность: в выборке ни одна из независимых переменных не постоянна, нет точной линейной зависимости между независимыми переменными (например, одна переменная не может быть постоянной, умноженной на другую)

3. Оценки несмещенные, менее эффективные, состоятельные.

Если мультиколлинеарность в модели выявлена, можно использовать следующие способы коррекции:

1) включить в модель другие переменные, которые не коррелируют друг с другом так сильно;

2) не менять модель (поскольку оценки несмещенные и состоятельные);

3) включить больше наблюдений в модель;

4) исключить одну из переменных и оставить только самые значимые;

5) изменить спецификацию модели.

Иногда количество наблюдений слишком велико, и исследователь не знает точно, какие переменные влияют на независимую переменную. Лучшим решением этой проблемы является выборочная процедура.

Предположим, что мы имеем p объясняющих переменных:

$$x_1, x_2, \dots, x_p.$$

Теоретически можем построить следующую модель:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon.$$

Но предположим, что число p слишком велико, и мы хотим знать, какие переменные должны быть включены в модель.

Алгоритм выборочной процедуры следующий:

1. Выполнить p парных регрессий y на x_1, x_2, \dots, x_p и после этого выбрать переменную, для которой коэффициент детерминации наибольший. На этом этапе мы находим наиболее информативные объясняющие переменные.

2. Выполнить $\frac{p \cdot (p-1)}{2}$ регрессий, каждый раз включая две

из p переменных, и, наконец, выбрать такую модель, которая пока-

зывает наибольшее значение коэффициента детерминации. Итак, теперь данная модель имеет две наиболее информативные переменные. **П р и м е ч а н и е:** переменная, выбранная на первом этапе, может быть исключена из этих переменных.

3. Найти три наиболее информативных переменных, выполняя $\frac{p \cdot (p-1)(p-2)}{6}$ регрессий.

4. Повторять этот алгоритм до тех пор, пока все переменные не будут включены в модель.

5. После всех предыдущих шагов нужно выбрать модель, для которой значение R_{\min}^2 наибольшее:

$$R_{\min}^2 = R_{adj}^2 - 2 \sqrt{\frac{2 \cdot k \cdot (N - K - 1)}{(n-1)(N^2 - 1)}} \cdot (1 - R^2(k)),$$

где k — количество включенных переменных; N — количество наблюдений; $R^2(k)$ — коэффициент детерминации, полученный на этапе k .

6. Модель, которая имеет самый высокий коэффициент детерминации, является наиболее информативной, и исследователь должен выбрать ее для дальнейшего анализа.

П р и м е ч а н и е. Этот метод является очень эффективным, но его недостаток в том, что слишком трудно выбрать такое огромное количество моделей. Другими словами, например, если у нас есть 20 объясняющих переменных, то мы должны запустить $2^{20} = 1\,048\,575$ регрессий. Вот почему в реальности применяется более простой метод.

Алгоритм более простой выборочной процедуры:

1. Выполнить p парных регрессий y на x_1, x_2, \dots, x_p и после этого выбрать переменную, для которой коэффициент детерминации наибольший. На этом этапе находим наиболее информативную объясняющую переменную.

2. Выполнить $(p-1)$ регрессий, каждый раз включая переменную, которая была выбрана на первом этапе, и некоторую переменную среди остальных $(p-1)$ переменных, и, наконец, выбрать

ту модель, которая показывает наибольшее значение коэффициента детерминации. Итак, пара регрессоров на первом и втором этапе является наиболее информативной.

3. Выполнить $(p - 2)$ регрессий, каждый раз включая переменные, которые были выбраны на первом и втором шагах алгоритма, и одну из остальных $(p - 2)$ переменных. Следовательно, мы выбрали 3 наиболее информативные переменные, включенные в модель, с наибольшим значением коэффициента детерминации.

4. Повторять этот алгоритм до тех пор, пока все переменные не будут включены в модель.

5. После всех предыдущих шагов необходимо выбрать модель, для которой значение R_{\min}^2 наибольшее:

$$R_{\min}^2 = R_{adj}^2 - 2 \sqrt{\frac{2 \cdot k \cdot (N - K - 1)}{(n - 1)(N^2 - 1)}} \cdot (1 - R^2(k)),$$

где k — количество включенных переменных; N — количество наблюдений; $R^2(k)$ — коэффициент детерминации, полученный на этапе k .

Модель с наибольшим коэффициентом детерминации является наиболее информативной, и исследователь должен использовать ее в дальнейшем анализе.

Существует две точки зрения относительно порядка выборочной процедуры. Некоторые могут подумать, что на самом деле разумная и правильная регрессия — это такая регрессия, которая включает в себя все объясняющие переменные. И только в этом случае с использованием МНК возможно получить несмещенную и эффективную оценку коэффициентов регрессии.

Другая точка зрения заключается в следующем: выборочная процедура — это процесс выбора истинной модели из огромного количества вероятных моделей, которые могут быть построены с использованием всех объясняющих переменных. Оценки коэффициентов регрессии являются несмещенными только после выполнения выборочной процедуры. Данная точка зрения преобладает.

В случае, если объясняющие переменные являются случайными (стохастическими) переменными, вопрос о правильности модели не возникает. Здесь мы должны найти модель, которая содержит

ошибку прогнозирования на рациональном уровне при ограниченном количестве переменных.

Эндогенность

В некоторых случаях нереально или даже невозможно рассматривать объясняющие переменные в модели как заданные или экзогенные. В таких случаях можно утверждать, что некоторые объясняющие переменные коррелируют с ошибкой уравнений таким образом, что оценка по МНК является смещенной и несостоятельной. Существуют разные причины, по которым можно утверждать, что ошибки одновременно коррелируют с одной или несколькими объясняющими переменными, но общий аспект заключается в том, что линейная модель больше не соответствует условному математическому ожиданию или наилучшей линейной аппроксимации. Таким образом, если условие $E(\varepsilon_i, x_i) = 0$ (ошибки и объясняющие переменные одновременно не коррелируют) не соблюдается по какой-либо причине, мы больше не утверждаем, что оценка по МНК является несмещенной или состоятельной, и нам нужно найти новую технику для оценки. Итак, если у нас есть линейная регрессия

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon, \text{ где } j = 1, \dots, k,$$

и условие $E(\varepsilon | x_1, x_2, \dots, x_k) = 0$ не выполняется по крайней мере для одной j , можно сделать вывод, что в этом случае x_j является эндогенной переменной. Причины могут быть разными, но мы рассмотрим только некоторые из них, а именно ошибки спецификации, пропущенную переменную, ошибки измерения (в зависимой или объясняющей переменной) и одновременность.

Предположим, что вместо обычной задачи включения в модель нерелевантной переменной мы опускаем переменную, которая на самом деле принадлежит ей. Это часто называют проблемой исключения релевантной переменной или недооценкой модели.

Например, возьмем простую регрессию заработной платы:

$$wage = \beta_0 + \beta_1 educ + \beta_2 ability + \varepsilon.$$

В данной модели заработная плата зависит от уровня образования работника и его способностей (это может быть физическая сила или интеллект, производительность и другие параметры). В приведенном выше уравнении заработной платы ε включает все ненаблюдаемые факторы, которые влияют на заработную плату человека, в том числе такие факторы, как «часы работы» или «опыт». Таким образом, переменная «способность» является эндогенной, поскольку мы пропускаем переменную «опыт», которая влияет как на заработную плату, так и на способность, т. е. $\text{corr}(\text{ability}, \varepsilon) \neq 0$.

Это очень важно, потому что оценки будут смещенными и несостоятельными.

$$\text{wage} = \beta_0 + \beta_1 \text{educ} + \varepsilon\varepsilon,$$

$$\varepsilon\varepsilon = \beta_2 \text{ability} + \zeta,$$

$$\text{corr}(\text{educ}, \text{ability}) \neq 0, \text{corr}(\text{educ}, \varepsilon\varepsilon) \neq 0,$$

$$\text{ability} = d_0 + d_1 \text{educ} + \zeta,$$

$$E(\hat{\beta}_1) = E(\hat{\beta}_1) + E(\hat{\beta}_2) \tilde{\delta}_1 = \beta_1 + \beta_2 \tilde{\delta}_1,$$

$$\text{BIAS} = E(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1.$$

Но если модель удовлетворяет условиям Гаусса — Маркова 1–4, то свойство несмещенности должно выполняться: $E(\hat{\beta}_1) = \beta_1$. Таким образом, пропуск релевантной переменной приводит к смещению МНК-оценок. Если ошибка ε коррелирует с educ , т. е. $\text{cov}(\text{educ}, \varepsilon) \neq 0$, то появляется несостоятельность в $\hat{\beta}_1$ (асимптотическое смещение):

$$\text{plim}(\hat{\beta}_1) - \beta_1 = \frac{\text{cov}(\text{educ}, \varepsilon)}{\text{var}(\text{educ})}.$$

Получение смещения, вызванного отсутствием важной переменной, является примером анализа ошибки спецификации. Формально

мы могли бы представить случай, когда верная модель совокупности имеет две объясняющие переменные и ошибку:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

Мы знаем, что

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N (x_{i1} - \bar{x}_1) y_i}{\sum_{i=1}^N (x_{i1} - \bar{x}_1)^2}.$$

Следующий шаг — самый важный. Поскольку наша регрессия является истинной моделью, мы запишем y для каждого наблюдения i как $y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$, теперь вставим данное уравнение в формулу для оценки β_1 и перепишем ее (для удобства не будем писать зна-

менатель $\sum_{i=1}^N (x_{i1} - \bar{x}_1)^2$ в полученном выражении):

$$\begin{aligned} & \sum_{i=1}^N (x_{i1} - \bar{x}_1) (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i) = \\ & = \beta_1 \sum_{i=1}^N (x_{i1} - \bar{x}_1)^2 + \beta_2 \sum_{i=1}^N (x_{i1} - \bar{x}_1) x_{i2} + \sum_{i=1}^N (x_{i1} - \bar{x}_1) u_i. \end{aligned}$$

Если мы разделим это уравнение на $\sum_{i=1}^N (x_{i1} - \bar{x}_1)^2$, примем, что

ожидание обусловлено значениями независимых переменных, и используем условие $E(u_i) = 0$, то получим

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\sum_{i=1}^N (x_{i1} - \bar{x}_1) x_{i2}}{\sum_{i=1}^N (x_{i1} - \bar{x}_1)^2}$$

Таким образом, $\hat{\beta}_1$ действительно смещена. Отношение, умноженное на β_2 , является просто коэффициентом наклона от регрессии x_2 на x_1 , с использованием нашей выборки из независимых переменных, его можно записать как

$$\hat{x}_2 = \hat{\delta}_0 + \hat{\delta}_1 x_1.$$

Поскольку мы обуславливаем значения выборки обеих независимых переменных, она не случайна здесь: $E(\hat{\beta}_1) = \beta_1 + \beta_2 \hat{\delta}_1$, таким образом, $E(\hat{\beta}_1) - \beta_1 = \beta_2 \hat{\delta}_1$.

Последние уравнения называются смещениями пропущенных переменных. Если нужна несмещенная оценка $\hat{\beta}_1$, мы должны приравнять $\beta_2 = 0$, и x_2 не будет участвовать в верной модели. Предположим, что нам нужно приравнять $\hat{\delta}_1 = 0$. $\hat{\delta}_1$ – выборочная ковариация между x_1 и x_2 по выборочной дисперсии x_1 , $\hat{\delta}_1 = 0$ тогда и только тогда, когда x_1 и x_2 не коррелированы в выборке. Таким образом, мы имеем важный вывод, что если x_1 и x_2 не коррелированы в выборке, тогда $\hat{\beta}_1$ не смещена.

В табл. 4 обобщены направления смещения, но размер смещения также очень важен для результатов оценки. Например, если отдача от образования у населения составляет 8,6 %, а смещение в МНК-оценке — 0,1 %, то мы не будем очень обеспокоены. С другой стороны, смещение порядка трех процентов считается гораздо более серьезным. Размер смещения определяется размерами β_2 и $\hat{\delta}_1$.

Таблица 4

Факторы, влияющие на смещение в случае пропущенной переменной

Фактор	$\text{corr}(x_1, x_2) > 0$	$\text{corr}(x_1, x_2) < 0$
$\beta_2 > 0$	Положительное смещение	Отрицательное смещение
$\beta_2 < 0$	Отрицательное смещение	Положительное смещение

Таким образом, исключение релевантной переменной приводит к несостоятельным МНК-оценкам. Подведем итоги. Если мы не включили переменную, но данная переменная релевантна или коррелирует с другой объясненной переменной, то, значит, мы пропустили переменную, которая приводит к проблеме эндогенности.

Следующая потенциальная причина эндогенности — *ошибки спецификации*. Это особый случай, когда пропущенная переменная является функцией объясняющей переменной в модели. Другими словами, мы сталкиваемся с этой проблемой, если модель регрессии имеет неправильную функциональную форму, а переменные связаны друг с другом неправильным образом. В целом следует отметить, что даже правильно заданная на данном этапе эконо-

трическая модель должна постоянно совершенствоваться на основе меняющихся условий функционирования экономических систем. Улучшение модели для лучшего соответствия экспериментальным данным может привести к изменению функциональной формы отношений между переменными и множеством факторов-аргументов.

Ошибки спецификации функциональной формы модели могут иметь серьезные последствия. Положительный момент в том, что по определению у нас есть сведения о всех необходимых переменных для получения функциональной формы, которая хорошо соответствует данным. Эта проблема может быть противопоставлена проблеме, рассматриваемой в следующих разделах, где ключевая переменная пропущена (по ней мы не можем собрать данные). Уже существует мощный инструмент для обнаружения неправильной функциональной формы: F -тест на совместные ограничения на исключение. Часто имеет смысл добавить квадратичные условия каких-либо значимых переменных в модели и провести совместный тест на значимость. Если дополнительные квадратичные члены значительны, то их можно добавить к модели (за счет усложнения интерпретации модели). Однако значимые квадратичные члены могут обусловить другие проблемы функциональной формы, например, использование уровня переменной, когда логарифм более уместен, или наоборот. Трудно определить точную причину того, что существует ошибка функциональной формы. К счастью, во многих случаях использование логарифмов определенных переменных и добавление квадратичных членов достаточно для обнаружения многих важных нелинейных отношений в экономике.

Для ясности: мы исследуем линейную регрессию $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$, но в реальной жизни, с учетом всех факторов, которые мы имеем для анализа, мы исследуем регрессии $y = \beta_0 + \beta_1 x_1 + \beta_2^2 x_2 + \dots + \beta_k x_k + u$.

Перечислим последствия такой ошибки:

- 1) смещение оценок $\beta_0, \beta_1, \beta_2$;
- 2) смещение зависит от размера β_2 и корреляции между x_1, x_2, \dots, x_k .

К ошибкам в спецификации приводит пропуск значительных факторов, т. е. использование обычной линейной модели вместо

множественной. Например, спрос на конкретный продукт может определяться не только ценой продукта, но и подушевым доходом потребителя. Для случая парной (одномерной) модели регрессии выбор функциональной формы обычно делается с учетом положения точек наблюдений в поле корреляции. Для оцененной регрессии оценка функциональной формы может быть выполнена путем анализа коэффициента детерминации и средней погрешности аппроксимации. Для множественной регрессии ситуация более неоднозначна, поскольку графическое представление статистических данных в этом случае невозможно. Здесь детальный анализ адекватности функциональной формы модели и ошибки спецификации может осуществляться на основе специальных тестов.

Существует возможность обнаружить ошибку спецификации с помощью F -теста или *RESET-теста* (тест регрессии на ошибку спецификации). Для начала нужно добавить квадратичные члены каких-либо значимых переменных и выполнить F -тест на совместную значимость и тогда, если дополнительные члены являются значимыми, добавить их в модель.

RESET-тест:

1. Оценить зависимую переменную следующей регрессии: $y = \beta_0 + \beta_1 x_1 + \dots + \varepsilon$, получить \hat{y}^2 .
2. Добавить \hat{y}^2 в регрессию $y = \beta_0 + \beta_1 x_1 + \dots + \beta_N \hat{y}^2 + \varepsilon$ и оценить значение β_N .

Еще одна иллюстрация: МНК-оценка является несостоятельной тогда, когда независимые переменные имеют ошибки измерения. Предположим, что мы имеем регрессионную модель $y^* = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$ и y^* — истинное значение зависимой переменной (например, ежегодные семейные сбережения), мы также имеем $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$, где y — наблюдаемое значение y^* и $y \neq y^*$ (например, годовые семейные сбережения), тогда $e_0 = y - y^*$ является точной погрешностью измерения. Добавим $y^* = y - e_0$ в истинную регрессию:

$$y - e_0 = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon,$$

ошибка: $v = \varepsilon + e_0$ и $E(v) = E(\varepsilon) + E(e_0)$.

Если ошибка измерения не коррелирует с независимыми переменными, то МНК-оценка имеет хорошие свойства. Последствиями ошибки этого типа являются смещенные и непоследовательные оценки, если $E(e_0) \neq 0$, $E(e_0 | x_1, x_2, \dots, x_k) \neq 0$, даже если $E(\varepsilon) = 0$ и $E(\varepsilon | x_1, x_2, \dots, x_k) = 0$. Также менее эффективны оценки, когда

$$\text{cov}(\varepsilon, e_0) = \sigma_\varepsilon^2 + \sigma_{e_0}^2 > \sigma_\varepsilon^2.$$

Ошибка измерения может быть обнаружена в объясняющих переменных:

$$y = \beta_0 + \beta_1 x_1^* + \varepsilon,$$

где x^* — истинное значение переменной и x — наблюдаемое значение x^* , ошибка измерения: $e_1 = x_1 - x^*$, тогда $y = \beta_0 + \beta_1 x_1 + (\varepsilon - \beta_1 e_1)$ при условии $\text{cov}(x_1^*, e_1) = 0$ и $x_1^* = x_1 - e_1$, $\text{cov}(x_1, e_1) = E(x_1 e_1) = E(x_1^* e_1) + E(e_1^2) = 0 + \sigma_{e_1}^2 = \sigma_{e_1}^2$:

$$\text{cov}(x_1, \varepsilon - \beta_1 e_1) = -\beta_1 \text{cov}(x_1, e_1) = -\beta_1 \sigma_{e_1}^2.$$

В данном случае МНК приводит к смещенным и несостоятельным оценкам.

Другой важной формой эндогенности объясняющих переменных является *одновременность*. Данная проблема возникает, когда одна или несколько независимых переменных определяются совместно с зависимой переменной, как правило, с помощью механизма равновесия (как мы увидим позже). В этой главе изучаются методы оценки моделей простых одновременных уравнений (SEM). Главным методом оценки моделей одновременных уравнений является метод инструментальных переменных (IV метод). Таким образом, решение проблемы одновременности по существу совпадает с решением IV для проблемы пропущенных переменных и ошибок измерения. Однако разработка и интерпретация SEM сложна.

Если мы рассмотрим уравнение, в котором одна или более независимых переменных совместно определены переменной с левой стороны, МНК-оценка будет давать, как правило, несостоятельные оценки поведенческих параметров в этом уравнении.

Проблема одновременности возникает, когда модель содержит поведенческие параметры. Если измерить причинно-следственные связи изменений в объясняющих переменных, то одна или несколько из этих объясняющих переменных совместно будут определяться переменной из левой части. Возьмем кейнсианскую функцию потребления:

$$C_t = \beta_1 + \beta_2 Y_t + \varepsilon,$$

где C_t отражает реальное потребление на душу населения в стране, а Y_t является реальным доходом на душу населения. Интерпретируем коэффициент β_2 как предельную склонность к потреблению ($0 < \beta_2 < 1$). Это значит, что β_2 имеет причинную интерпретацию, отражающую влияние дохода на потребление: насколько больше люди потребляют, если их доход возрастает на одну единицу? Однако совокупный доход Y_t не является экзогенно заданным, поскольку он будет определяться как

$$Y_t = C_t + I_t,$$

где I_t — реальные инвестиции на душу населения. Это уравнение представляет собой уравнение определения для закрытой экономики без государственных расходов и показывает, что сумма совокупного потребления и совокупных инвестиций должна быть равна совокупному доходу.

Формально проблема одновременности может быть представлена как

$$y_1 = a_1 y_2 + b_1 z_1 + u_1,$$

$$y_2 = a_2 y_1 + b_2 z_2 + u_2,$$

$$\text{cov}(z_1, u_1) = 0, \text{cov}(z_1, u_2) = 0,$$

$$\text{cov}(z_2, u_1) = 0, \text{cov}(z_2, u_2) = 0,$$

поэтому z_1 и z_2 являются экзогенными.

Вставим уравнение для y_1 в уравнение y_2 :

$$y_2 = a_2 (a_1 y_2 + b_1 z_1 + u_1) + b_2 z_2 + u_2,$$

$$(1 - a_2 a_1) y_2 = a_2 b_1 z_1 + b_2 z_2 + a_2 u_1 + u_2, \text{ если } a_2 a_1 \neq 1,$$

и получим

$$y_2 = p_{21} z_1 + p_{22} z_2 + v_2.$$

Последнее уравнение отражает y_2 через экзогенные переменные и ошибки, это приведенная форма для y_2 в контексте оценки инструментальных переменных. Параметры p_{21} и p_{22} называются приведенными параметрами формы; заметим, что они являются нелинейными функциями структурных параметров.

Если разделим предыдущее уравнение на $(1 - a_2 a_1)$, получим

$$p_{21} = \frac{a_2 b_1}{(1 - a_2 a_1)}, \quad p_{22} = \frac{b_2}{(1 - a_2 a_1)}, \quad v_2 = \frac{a_2 u_1 + u_2}{(1 - a_2 a_1)}.$$

Приведенная ошибка формы v_2 является линейной функцией членов структурной ошибки u_1 и u_2 . Поскольку u_1 и u_2 (каждая) не коррелирует с z_1 и z_2 , v_2 также не коррелирует с z_1 и z_2 . Поэтому мы можем последовательно оценить p_{21} и p_{22} по МНК, используя для оценки наименьших квадратов два этапа (более подробно в следующем разделе). Кроме того, снижение параметра формы иногда представляет непосредственный интерес, хотя мы сосредоточились здесь на оценке уравнения.

Когда $a_2 = 0$, u_1 и u_2 не коррелируют, y_2 и u_1 также не коррелируют. Довольно строгие ограничения: если $a_2 = 0$, то y_2 не определяется одновременно с u_1 . Если мы добавим нулевую корреляцию между u_1 и u_2 , то исключим пропущенные переменные или ошибку измерения в u_1 , которая коррелирует с y_2 . Мы не должны удивляться тому, что оценка по МНК уравнения $y_1 = a_1 y_2 + b_1 z_1 + u_1$ работает в этом случае. Когда y_2 коррелирует с u_1 из-за одновременности, можно отметить, что МНК подвергается смещению одновременности.

Когда $\text{cov}(y_2, u_1) \neq 0$, можно сделать вывод, что МНК будет давать смещенные и несостоятельные оценки.

Каждая проблема эндогенности имеет свое собственное решение. Например, как отмечалось ранее, если мы имеем дело с проблемой функциональной формы, мы должны использовать тест на ошибку спецификации формы, также мы можем добавить логарифмический или квадратичный член формы в модель. Скорее всего, это будет решение проблемы, и мы получим репрезентативную регрессию.

Когда мы имеем дело с пропущенной переменной, мы должны использовать прокси-переменную (замещающую переменную). Однако если данные для опущенной переменной существуют, включите ее в регрессию. Прокси-переменная — это такая переменная, которая связана с ненаблюдаемой переменной.

Давайте рассмотрим пример получения высокого балла на экзамене и найдем, какие факторы повлияют больше всего на вероятность успешных результатов. Предположим, есть регрессия следующего вида:

$$\log(A) = \beta_0 + \beta_1 X_h + \beta_2 X_e + \beta_3 X_a + \varepsilon,$$

где A — возможная оценка; X_h — количество учебных часов; X_e — уровень образования; X_a — логическая способность ученика. Мы не можем напрямую измерять способность, поэтому нам нужно найти прокси-переменную. Пусть ею будет коэффициент IQ. Итак, $\text{cov}(IQ, X_a) \neq 0$, и мы можем включить ее в регрессионную модель, а затем оценить ее.

Иногда очень трудно найти точные сочетания переменных, в этом случае лаг может играть роль прокси-переменной.

С точки зрения оценки модели это было бы

$$\log(A) = \beta_0 + \beta_1 X_h + \beta_2 X_e + \beta_3 X_{A-1} + \varepsilon,$$

где X_{A-1} означает предыдущую оценку студента, полученную в предыдущем периоде (может быть, десять лет назад или два года назад, и мы можем включить более чем один лаг). В этих условиях мы ожидаем, что β_3 больше нуля. В другом случае мы рискуем получить смещенные оценки и неэффективную модель. Другой способ смягчить проблему пропущенных переменных — использовать оценку первых разностей или фиксированных эффектов.

Следующий способ решения проблемы эндогенной переменной — включить *инструментальную переменную (IV)*. IV — это переменная, которая коррелирует с проблемной переменной в модели, но никак не коррелирует с ошибкой. Итак, если говорить о стандартной модели регрессии, мы можем формально представить этот факт:

$$y = \alpha + \beta x + \varepsilon.$$

Предположим, что Z — инструментальная переменная, тогда, согласно определению, $\text{cov}(x, z) \neq 0$ и $\text{cov}(z, \varepsilon) = 0$. Выбрав данную IV, мы запускаем *тест на слабую IV (weak IV test)*, чтобы удостовериться, что z может быть включена в модель. Существует новая регрессия для x :

$$x = \pi_0 + \pi_1 z + \varepsilon x = \pi_0 + \pi_1 z + \varepsilon,$$

со стандартными двухсторонними гипотезами:

$$H_0 : \pi_1 = 0,$$

$$H_1 : \pi_1 \neq 0.$$

Если H_0 отклоняется, то у нас есть нормальная прокси-переменная, которую можно включать в модель.

Теперь мы формализуем, что прокси-переменная является эффективным инструментом в решении эндогенной проблемы.

В регрессионной модели $y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k + \varepsilon$, в матричном виде $y = \beta x + \varepsilon$ и получим

$$\begin{aligned} \hat{\beta}_{IV} &= (Z^T X)^{-1} Z^T Y, \\ \hat{\beta}_{IV} &= \frac{\sum_{i=1}^N Z_i Y_i - \bar{Z} \bar{Y}}{\sum_{i=1}^N Z_i X_i - \bar{Z} \bar{X}} = \frac{\widehat{\text{cov}}(Z, Y)}{\widehat{\text{cov}}(Z, X)} = \frac{\widehat{\text{cov}}(Z, \beta x + \varepsilon)}{\widehat{\text{cov}}(Z, X)} = \beta + \frac{\widehat{\text{cov}}(Z, \varepsilon)}{\widehat{\text{cov}}(Z, X)}, \\ \underset{n \rightarrow \infty}{\text{plim}} \hat{\beta}_{IV} &= \beta + \frac{\underset{n \rightarrow \infty}{\text{plim}} \widehat{\text{cov}}(Z, \varepsilon)}{\underset{n \rightarrow \infty}{\text{plim}} \widehat{\text{cov}}(Z, X)} = \beta + \frac{\widehat{\text{cov}}(Z, \varepsilon)}{\widehat{\text{cov}}(Z, X)} = \beta. \end{aligned}$$

Тогда $\widehat{\text{cov}}(Z, \varepsilon) = 0$, если Z — эффективная прокси-переменная.

$$\begin{aligned}\text{var}(\hat{\beta}_{IV}) &= \text{var}\left(\left((Z^T X)^{-1} Z^T Y\right)\right) = \\ &= \text{var}\left(\left((Z^T X)^{-1} Z^T (\beta x + \varepsilon)\right)\right) = \text{var}\left(\left((Z^T X)^{-1} Z^T \beta\right) + (Z^T X)^{-1} Z^T \varepsilon\right) = \\ &= \text{var}\left(\left((Z^T X)^{-1} Z^T \varepsilon\right)\right) = (Z^T X)^{-1} Z^T \text{var}\left(\left(Z^T X\right)^{-1}\right) = \sigma^2 (Z^T X)^{-1} Z^T Z (Z^T X)^{-1},\end{aligned}$$

где σ^2 определяется по формуле

$$\sigma^2 = \frac{ESS}{N - K - 1} = \frac{\sum_{i=1}^N (Y_i - X^T \hat{\beta}_{IV})^2}{N - K - 1}.$$

Механизм работы IV:

$$y = \alpha + \beta x + \varepsilon,$$

$$\text{cov}(z, \varepsilon) = 0,$$

$$\text{cov}(z, x) \neq 0,$$

$$\text{cov}(z, y) = \beta \text{cov}(z, x) + \text{cov}(z, \varepsilon),$$

$$\hat{\beta} = \frac{\text{cov}(z, y)}{\text{cov}(z, x)},$$

$\text{cov}(z, x)$ должна быть высокой, в противном случае оценки будут еще более смещенными при МНК.

Существует еще один способ борьбы с эндогенностью — оценка по *двухшаговому методу наименьших квадратов (2SLS)*. IV и 2SLS могут использоваться для получения состоятельных оценок в случае пропущенных переменных, также для решения проблемы ошибок в переменных и могут помочь оценить модели одновременных уравнений.

Для дальнейшего анализа мы вводим следующие структурные и приведенные формы уравнений (структурная форма представляет собой уравнение с эндогенными и экзогенными переменными,

а приведенная форма будет содержать инструментальные переменные плюс экзогенные независимые переменные):

$Y = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$ — структурное уравнение формы,

$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2$, $\pi_2 \neq 0$ и $\pi_3 \neq 0$ — приведенное уравнение формы.

Когда выполним последнюю регрессию и получим \hat{y}_2 , будем использовать его в качестве инструментальной переменной для первой регрессии на y_2 . Таким образом, оценка \hat{y}_2 , используемая в качестве инструментальной для y_2 в первой регрессии y , называется двухшаговой оценкой наименьших квадратов.

Сейчас мы имеем регрессию $y = \alpha_1 y_2 + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$, где y_2 — эндогенная переменная. Предположим, что в данном случае m инструментальных переменных. Инструменты: $z = (1, x_1, \dots, x_k, z_1, \dots, z_m)$.

Запишем приведенную форму для уравнения y_2 :

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \dots \pi_k z_k + \pi_{k+1} z_1 + \pi_{k+m} z_m + \varepsilon,$$

тогда $y_2 = \hat{y}_2 + \varepsilon$.

Следующий шаг очень важный, потому что \hat{y}_2 не коррелирует с u . Таким образом, можно сказать, что, оценивая y_2 со всеми экзогенными переменными, мы разделили его на две части: одна коррелирует с u , а другая — нет.

Можно записать \hat{y}_2 в матричном виде как $Z(Z^T Z)^{-1} Z^T y_2$. Мы также можем ввести $\hat{x} = Z(Z^T Z)^{-1} Z^T x$ в матричном виде.

Мы используем \hat{x} в качестве инструментов для x и применим оценку IV:

$$\begin{aligned} \hat{\beta}_{2SLS} &= (\hat{X}^T X)^{-1} \hat{X}^T Y = \left(X^T Z (Z^T Z)^{-1} Z^T X \right)^{-1} X^T Z (Z^T Z)^{-1} Z^T Y = \\ &= \left(X Z (Z^T Z)^{-1} Z^T X \right)^{-1} X^T Z (Z^T Z)^{-1} Z^T Y. \end{aligned}$$

В простой форме можно также переписать как

$$\hat{\beta}_{2SLS} = (\hat{X}^T \hat{X})^{-1} \hat{X}^T Y.$$

Мы можем просто оценить 2SLS-оценки за один шаг, используя X и Z . Но в обычном случае для оценки 2SLS-оценки используется двухшаговая процедура.

В следующей главе описывается шаг номер один (регрессия первого шага) и второй шаг (регрессия второго шага).

Мы выяснили, что каждая проблема эндогенности имеет свое собственное решение. Но прежде чем решать проблемы, мы должны их обнаружить. Итак, как можно обнаружить эндогенность?

2SLS менее эффективен, чем МНК, если объясняющие переменные являются экзогенными. Необходим тест на эндогенность, чтобы увидеть, требуется ли вообще 2SLS.

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \varepsilon,$$

z_1 и z_2 — экзогенные, также есть z_3 и z_4 , которые не записаны в регрессии

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \varepsilon.$$

Если $\text{corr}(y_2, \varepsilon) = 0$, то мы можем применить МНК для

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \varepsilon.$$

Наиболее эффективным способом обнаружения эндогенности является тест Хаусмана, сравнивающий оценки МНК и 2SLS и определение того, являются ли различия статистически значимыми. Предположим, что мы хотим проанализировать y_i с помощью x_{1i} и ряда контролов x_{ji} , где $j \in 1, 2, \dots, k$. Предположим, x_{1i} — эндогенная. Существует n инструментов Z_{ni} для прогнозирования x_{1i} . 2SLS-модель принимает вид:

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \varepsilon,$$

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \pi_4 z_4 + v,$$

после оценки мы получим \hat{v} $\text{corr}(y_2, \varepsilon) = 0$ тогда и только тогда, когда $\text{corr}(v, \varepsilon) = 0$.

Включим \hat{v} в исходную регрессию и выполним t -тест на значимость (против двусторонней альтернативы).

Если H_0 отклоняется, то можно сделать вывод, что y_2 — эндогенная, потому что $\text{corr}(v, \varepsilon) \neq 0$.

МНК-оценки (кроме d_1) для модели $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \delta_1 \hat{v} + \varepsilon$ идентичны оценкам 2SLS для модели $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \varepsilon$, поэтому мы можем проверить, выполнили ли мы правильную регрессию при тестировании на эндогенность.

Следующая проблема, которую можно протестировать, — *сверхидентифицирующие ограничения* (*overidentifying restrictions*). Тогда и только тогда, если уравнение сверхидентифицировано, мы можем проверить, не исключены ли инструменты, независимые от процесса ошибки. Этот тест всегда должен выполняться, когда это возможно, поскольку позволяет оценить валидность инструментов. Тест на сверхидентифицирующие ограничения возвращает остатки с IV или 2SLS регрессии по всем инструментам в Z . При нулевой гипотезе все инструменты не связаны с u , тестируется большая выборка $\chi^2(r)$ распределения, где r — количество сверхидентифицирующих ограничений.

Прежде всего, оценка IV должна удовлетворять условию $\text{cov}(z, \varepsilon) = 0$, и если оно выполняется, то можно построить приведенное уравнение формы и выполнить t -тест. Когда IV удовлетворяет условию $\text{cov}(z, x) \neq 0$, мы также можем проверить регрессию, используя t -тест, но только тогда, когда у нас есть не более чем одна инструментальная переменная. В общем, число инструментальных переменных имеет ключевое значение. Предположим, что NIV — это число инструментальных переменных, а NEX — количество экзогенных переменных.

Существует три случая, определяющих способность оценивать:

- 1) $NIV < NEX$ — в этом случае мы имеем недооценку и не можем оценить модель;
- 2) $NIV > NEX$ — имеется сверхидентификация, поэтому мы можем напрямую протестировать на сверхидентифицирующие ограничения;
- 3) $NIV = NEX$ — не можем протестировать на сверхидентифицирующие ограничения.

Таким образом, в случае 2 мы можем применить тестирование на сверхидентифицирующие ограничения. Ниже приведен алгоритм оценки.

Во-первых, мы формулируем стандартные гипотезы:

H_0 : все IV не коррелируют с ε ;

H_1 : по крайней мере некоторые из IV не являются экзогенными.

Затем мы оценим структурное уравнение с помощью 2SLS и получим остатки $\hat{\varepsilon}$, после чего регрессируем $\hat{\varepsilon}$ по всем экзогенным переменным и получим R^2 .

Далее мы используем известную тестовую статистику nR^2 и сравниваем с правилом отклонения: $nR^2 > \chi_{q(cr)}^2$, где q — количество сверхидентифицирующих ограничений, и если $nR^2 > \chi_{q(cr)}^2$, отвергаем H_0 .

Если модель содержит некоторые ошибки, связанные с функциональной формой, исследователь должен справиться с ними, используя три лучших метода: F -тест, $RESET$ -тест (regression specification error test) и тест против невложенных альтернатив.

F -тест может показать объясняющую силу оригинальной модели и модели с функциональными изменениями. Благодаря знаниям об объясняющей силе становится понятным, нужны ли функциональные модификации или нет. Алгоритм F -теста был описан ранее. Вкратце, если F -тест показывает, что объясняющая сила модели с функциональными изменениями (например, с квадратичными членами) выше, чем объясняющая сила исходной модели, то существует ошибка функциональной формы и необходимо каким-либо образом изменить модель.

Для $RESET$ -теста генерируем новую регрессию, а затем оцениваем ее с использованием F -теста. Так, если существует регрессия $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$, но не известна точная лучшая форма спецификации, и только подозревая, что что-то не так, мы создаем новую регрессию, подобную этой: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u$. Если модель правильно указана, то $H_0: \delta_1 = 0, \delta = 0$ и H_1 : в модели есть функциональная проблема. Затем мы проверяем F -тест. $RESET$ -тест также может быть эффективным в случае двух и более моделей, тогда мы должны сделать $RESET$ для каждой

модели и выбрать тот, у которого большая статистика *RESET* (или с меньшим информационным критерием).

Третий метод — это тест против невложенных альтернатив. Тесты невложенных гипотез дают возможность проверить спецификацию эконометрической модели на основании доказательств, предоставленных одной или несколькими невложенными альтернативами. Покажем, как это работает. Предположим, что функция регрессии

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

и регрессия, которая может увеличить объясняющую силу:

$$y = \beta_0 + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \varepsilon.$$

Чтобы проверить, какая модель лучше, мы строим новую модель, а затем проверяем следующие гипотезы:

$$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 \ln(x_1) + \gamma_4 \ln(x_2) + \varepsilon,$$

$H_0: \gamma_1 = 0, \gamma_2 = 0$, как тест модели с логарифмом,

$H_0: \gamma_3 = 0, \gamma_4 = 0$, как тест исходной модели.

Этот подход был предложен Мизоном и Ричардом. Другой способ — использовать тест Дэвидсона — Мак-Киннона:

На первом этапе мы оцениваем (с помощью МНК) функцию регрессии с логарифмом и получаем \hat{y} . На втором этапе включаем \hat{y} в исходную модель и выполняем t -тест против двухсторонней альтернативы. Если мы получаем значимую t -статистику, мы отклоняем исходную модель и говорим, что она имеет функциональное отклонение формы.

Тесты против невложенных альтернатив имеют некоторые ограничения:

- 1) отвержение обеих моделей — если у нас есть такие результаты, мы должны точно изменить структуру модели;
- 2) включены обе модели — в этом случае мы сравниваем модели, находя наибольший R^2 ;
- 3) отказ от одной модели не означает, что другая модель верна;
- 4) трудно выполнять тесты, если модели имеют разные зависимые переменные.

Таким образом, проблема функциональной ошибки спецификации не всегда легко решается. Чтобы построить правильную форму модели, мы должны знать полную информацию об исследуемых объектах и всех функциональных связях.

Как правило, нет ничего плохого в том, чтобы включить в модель переменные, которые коррелированы. Например, в уравнение индивидуальной заработной платы можно включать как возраст, так и опыт, хотя можно ожидать, что пожилые люди в среднем имеют больше опыта. Однако если корреляция между двумя переменными слишком велика, это может привести к некоторым проблемам. Технически проблема заключается в том, что матрица $X^T X$ почти необратима. Это может привести к ненадежным оценкам с высокими стандартными ошибками и неожиданным знаком или величиной. Интуитивно проблема также очевидна. Если возраст и опыт сильно коррелированы, то может быть трудно идентифицировать индивидуальное воздействие этих двух переменных. В таком случае большое количество наблюдений с достаточным изменением как возраста, так и опыта может помочь нам получить разумные ответы. Если это не так, и мы получаем плохие оценки (например, t -тесты показывают, что ни возраст, ни опыт не являются индивидуально значимыми), мы можем сделать вывод о том, что в выборке недостаточно информации для выявления эффектов, которые мы хотели бы идентифицировать. В общем случае термин «мультиколлинеарность» используется для описания проблемы, когда приближенная линейная зависимость между объясняющими переменными приводит к ненадежным оценкам регрессии. Это приблизительное соотношение не ограничивается двумя переменными, оно может включать даже все регрессоры. Например, в уравнении заработной платы проблемы могут усугубляться, если мы включаем годы обучения в дополнение к возрасту и многолетнему опыту. В крайнем случае одна объясняющая переменная представляет собой точную линейную комбинацию одной или нескольких других объясняющих переменных (включая точку пересечения). Это обычно называют полной мультиколлинеарностью, когда МНК-оценка не определяется однозначно из условий первого порядка задачи наименьших квадратов (матрица $X^T X$ необратима). Использование слишком

большого количества фиктивных переменных (каждая из которых ноль или один) является типичной причиной полной мультиколлинеарности.

Таким образом, мультиколлинеарность часто является проблемой выборки, но не общей проблемой совокупности. Коллинеарность означает существование линейной связи между переменными, а мультиколлинеарность означает очень сильную линейную связь между переменными. Поэтому, когда у нас есть регрессия,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad \text{и} \quad \hat{\beta}_{MKN} = (X^T X)^{-1} X^T Y,$$

а мультиколлинеарность — это ситуация, когда $\det(X^T X) = 0$, $\hat{\beta}_{MKN}$ формально существует.

Есть наиболее распространенные причины мультиколлинеарности:

1. Существует полная коллинеарность, когда $\det(X^T X) = 0$ и $\text{rank}(X) < k + 1$ в матричном виде.

$$X = \begin{pmatrix} 1 & X_{11} & X_{21} & X_{k1} \\ 1 & X_{12} & X_{22} & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1N} & X_{2N} & X_{kN} \end{pmatrix}$$

В матрице есть линейно зависимые столбцы. Это может быть ситуация, когда мы выбираем только женщин или только мужчин (например, регрессия для заработной платы). Это означает, что мы используем только одну характеристику фиктивной переменной. $Wage = \beta_0 + \beta_1 age + \beta_2 edu + \beta_3 sex + \varepsilon$, где age — возраст респондента, edu — уровень образования, sex — фиктивная переменная со значением 1 для женщин и 0 для мужчин.

$$Y = \begin{pmatrix} 1 & age_{11} & edu_{21} & 1 \\ 1 & age_{12} & edu_{22} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & age_{1N} & edu_{2N} & 1 \end{pmatrix} \quad Y = \begin{pmatrix} 1 & age_{11} & edu_{21} & 0 \\ 1 & age_{12} & edu_{22} & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & age_{1N} & edu_{2N} & 0 \end{pmatrix}$$

2. Переменные в модели имеют мощную связь:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon.$$

Например, x_1, x_2, x_3 являются составными частями чего-то единого и $x_1 + x_2 + x_3 = 100\%$. Это может быть состав ткани *cotton + lukra + polyamide* = 100% или *proteins + fats + carbohydrates* = 100% общих веществ. Включение данных переменных в регрессии по отдельности, скорее всего, станет причиной мультиколлинеарности.

3. Фиктивная ловушка. Это будет легко понять на следующем примере.

Предположим, что мы имеем регрессию, где

$$Y = \alpha + \beta S + \varepsilon$$

$$S = \begin{cases} 0, \text{ summer} \\ 1, \text{ autumn} \\ 2, \text{ winter} \\ 3, \text{ spring} \end{cases}$$

Существует фиктивная ловушка для фиктивной переменной:

$$d_0 = \begin{cases} 1, \text{ когда } \textit{summer} \\ 0, \text{ другой месяц} \end{cases} \quad d_2 = \begin{cases} 1, \text{ когда } \textit{autumn} \\ 0, \text{ другой месяц} \end{cases}$$

$$d_3 = \begin{cases} 1, \text{ когда } \textit{winter} \\ 0, \text{ другой месяц} \end{cases} \quad d_4 = \begin{cases} 1, \text{ когда } \textit{spring} \\ 0, \text{ другой месяц} \end{cases}$$

$$d_0 + d_2 + d_3 + d_4 = 1.$$

Пример исследования среднего спроса на мороженое:

$$y = \beta_0 + \beta_1 d_0 + \beta_2 d_2 + \beta_3 d_3 + \beta_4 d_4 + \varepsilon.$$

Если данные верны, модель никогда не может быть оценена. Одна из переменных должна быть удалена. Коэффициенты с фиктивными переменными показывают изменение переменных для каждого из уровней по сравнению с базовыми.

1. Мультиколлинеарность имеет технические последствия и сложность интерпретации результатов моделирования. Начнем с рассмотрения технических последствий.

Мультиколлинеарность искажает результаты. Когда мы делаем небольшое изменение в данных, значения коэффициентов резко изменяются.

$$\det(X^T X) = 0, \text{ так что } (X^T X)^{-1} = \frac{1}{\det} \begin{pmatrix} 1 & X_{11} & X_{21} & X_{k1} \\ 1 & X_{12} & X_{22} & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1N} & X_{2N} & X_{kN} \end{pmatrix}.$$

Таким образом, элементы этой матрицы не имеют ограничений и могут быть очень большими. Это опасно, поскольку элементы матрицы принимают участие в формировании дисперсии. $\text{var}(\hat{\beta}_i) = \sigma^2 a_{ij}$ увеличивается, невозможно становится оценить коэффициенты, они неправдоподобны.

2. Когда $\text{var}(\hat{\beta}_i) = \sigma^2 a_{ij}$ увеличивается, стандартная ошибка может быть раздута: $S_i = \sqrt{S^2 a_{ij}}$. Это является причиной низкой t -статистики. Другими словами, статистика показывает, что переменная незначительно влияет на зависимую переменную. В то же время она может оказать сильное воздействие. Такие результаты мы получаем только из-за мультиколлинеарности. То есть t -тест неэффективен при наличии мультиколлинеарности.

3. Результаты могут быть изменены. Небольшие изменения в модели и данных приводят к большим изменениям в интерпретации.

Поскольку феномен Стьюдента формируется за счет тесной статистической связи, можно попробовать найти его. Способы обнаружения мультиколлинеарности:

- построить матрицу парных коэффициентов. Если корреляция больше 0,8, то можно подозревать существование мультиколлинеарности;

- построить регрессию одной переменной на другую и обратить внимание на R^2 . Если его значение больше 0,8, то снова можно подозревать существование мультиколлинеарности.

Существует коэффициент вариации дисперсии для проверки этого предположения:

$$\text{vif}(j) = \frac{1}{1 - R_j^2}.$$

Также есть некоторые факторы, которые могли бы идентифицировать существование этого феномена:

- нереалистичные коэффициенты;
- высокая F -статистика;
- общая регрессия значима, но многих переменных нет;

Продемонстрируем влияние мультиколлинеарности на МНК-оценку. Предположим, что мы имеем следующую регрессию:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i.$$

Предположим следующие условия:

- 1) $\bar{y} = \bar{x}_1 = \bar{x}_2 = 0$,
- 2) $x_1 = 1$ и $x_2 = 1$,
- 3) коэффициент корреляции, который показывает, что связь между x_1 и x_2 равна r_{12} .

В этих условиях мы можем записать МНК-оценки как

$$V(b) = \frac{\sigma^2}{N} \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} = \frac{\sigma^2/N}{1 - r_{12}^2} \begin{pmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{pmatrix}.$$

Очевидно, что отклонения β_1 и β_2 возрастают, если возрастает абсолютное значение коэффициента корреляции между x_1 и x_2 . Из-за увеличенной дисперсии оценки по МНК t -статистика будет меньше.

Другим следствием мультиколлинеарности является то, что некоторые линейные комбинации параметров довольно точно оцениваются, а другие линейные комбинации являются неточными. Обычно, когда регрессоры положительно коррелированы, сумма коэффициентов регрессии может быть довольно точно определена, а разница нет. В приведенном выше примере мы имеем дисперсию $\beta_1 + \beta_2$ такую, что

$$V(\beta_1 + \beta_2) = \frac{\sigma^2/N}{1 - r_{12}^2} (2 - 2r_{12}) = 2 \frac{\sigma^2/N}{1 + r_{12}^2},$$

а также:

$$V(\beta_1 - \beta_2) = \frac{\sigma^2/N}{1 - r_{12}^2} (2 + 2r_{12}) = 2 \frac{\sigma^2/N}{1 - r_{12}^2}.$$

Поэтому если r_{12} близок к 1, дисперсия $\beta_1 - \beta_2$ во много раз превышает дисперсию $\beta_1 + \beta_2$.

Таким образом, высокие корреляции между объясняющими переменными (их линейными комбинациями) могут приводить к проблемам мультиколлинеарности. Если это происходит, один или несколько параметров, которые нас интересуют, оцениваются очень неточно. По сути, это означает, что наша выборка не предоставляет достаточной информации об этих параметрах. Чтобы облегчить проблему, мы вынуждены использовать больше информации, например, путем введения некоторых априорных ограничений на вектор параметров. Обычно это означает, что одна или несколько переменных пропущены из модели. Другим решением, которое обычно непрактично, является расширение размера выборки.

До сих пор были продемонстрированы основные проблемы мультиколлинеарности, и можно предположить, что есть некоторые способы борьбы с данной проблемой. Прежде всего, нужно включать переменные, которые коррелируют слабее. Также мы можем увеличить выборку и размер данных. Нужно уделять больше внимания выбору объясняющей переменной. Наиболее эффективный метод может помочь нам построить новые независимые переменные на основе базовых переменных.

ГЛАВА 3. СИСТЕМЫ РЕГРЕССИОННЫХ УРАВНЕНИЙ

Эконометрические модели часто пытаются объяснить группу взаимосвязанных переменных, например: уравнения спроса факторов и продуктов, инвестиционные функции для связанных фирм, уравнения заработной платы для соответствующих групп работников.

В этом случае необходимо использовать систему регрессионных уравнений. Система регрессионных уравнений — это совокупность эконометрических уравнений (часто линейных), определяющих взаимозависимость экономических переменных. Важная особенность системы регрессионных уравнений — появление одних и тех же переменных в правой и левой стороне различных уравнений в системе. Так, некоторые переменные могут быть зависимыми в некоторых уравнениях и в то же время они могут быть независимыми (объясняющими) переменными в других.

Системные оценки позволяют найти совместную оценку других несвязанных уравнений, которые связаны ковариационной структурой их ошибок (внешне несвязанной регрессии), а также уравнений, которые связаны между собой ограничениями коэффициента уравнений (например, суммирования и ограничения по симметрии теории спроса).

Внешне не связанные регрессии

Один из видов системы регрессионных уравнений — *внешне не связанные регрессии (SUR)*. Это система эконометрических уравнений; каждое из них является независимым уравнением с собственной зависимой переменной и объясняющими переменными. Важной особенностью данных уравнений является то, что, несмотря на кажущуюся бессвязность, их ошибки коррелируют друг с другом.

Предположим, что имеется следующая система уравнений:

$$\begin{cases} y_1 = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1k}x_k + \varepsilon_1 \\ y_2 = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2k}x_k + \varepsilon_2 \\ \dots \\ y_m = \beta_{m0} + \beta_{m1}x_1 + \beta_{m2}x_2 + \dots + \beta_{mk}x_k + \varepsilon_m \end{cases},$$

$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ коррелируют.

Существует предположение, что ошибка в каждом уравнении удовлетворяет условиям Гаусса — Маркова об отсутствии автокорреляции и гетероскедастичности. Но есть корреляция ошибок между уравнениями.

Пример. Модель, которая исследует характеристики автомобиля:

$$\begin{cases} price = \beta_0 + \beta_1 foreign + \beta_2 milage + \beta_3 displacement + \varepsilon_1 \\ weight = \gamma_0 + \gamma_1 foreign + \gamma_2 length + \varepsilon_2 \end{cases}$$

Если объясняющие переменные в SUR одинаковы во всех уравнениях, то коэффициенты такие же, как в уравнениях, оцененных отдельно по МНК.

Если коэффициенты ошибок SUR-уравнений коррелированы, то совместная оценка SUR более эффективна, чем отдельная оценка по МНК, даже если переменные одинаковы.

Модели одновременных уравнений

Наиболее широко используемыми на практике в эконометрических моделях являются модели одновременных уравнений (SEM). Рассмотрим следующую модель, описывающую равновесное количество рабочей силы (в часах) в некоторых секторах, например, в сельскохозяйственном секторе:

$$\begin{cases} h_d = \alpha_1 w + \beta_1 z_1 + \varepsilon_1 \\ h_s = \alpha_2 w + \beta_2 z_2 + \varepsilon_2 \end{cases},$$

где h_s — количество предлагаемых часов труда в аграрном секторе; h_d — часы труда, востребованные в аграрном секторе. Эти переменные зависят от w (ставки заработной платы) и z_1, z_2 (данные переменные отражают некоторые другие факторы).

Так, z_1 может быть ставкой заработной платы в каком-либо другом секторе, например, в производственном. Если заработная плата в обрабатывающей промышленности увеличится, люди перейдут в производственный сектор, и количество рабочих часов в сельскохозяйственном секторе сократится. Z_1 называется наблюдаемым сдвигом спроса, ε_1 — ненаблюдаемым сдвигом спроса.

Z_2 может обозначать площадь сельскохозяйственной земли. Если больше земли доступно, то необходимо больше рабочих, следовательно, спрос на рабочую силу возрастает. Аналогично z_2 называется наблюдаемым сдвигом предложения, а ε_2 — ненаблюдаемым сдвигом предложения. Спрос и предложение описывают совершенно разные отношения.

Соотношение труда и заработной платы определяется показателями h_d и h_s .

Равновесие:

$$h_s = h_d.$$

Предположим, что у нас есть данные по ряду стран. Мы могли бы определить равновесие предложения рабочей силы и заработной платы для каждой страны.

Спрос:

$$h_i = \alpha_1 w_i + \beta_1 z_{i1} + \varepsilon_{i1}.$$

Предложение:

$$h_i = \alpha_2 w_i + \beta_2 z_{i2} + \varepsilon_{i2},$$

где i — индекс страны.

Эти два уравнения образуют модель одновременных уравнений (SEM). Данные два уравнения называются структурными уравнениями. $\alpha_1, \beta_1, \alpha_2, \beta_2$ называются структурными параметрами.

В одновременных уравнениях модельные переменные h_i и w_i являются эндогенными переменными, так как они определяются

равновесием между двумя уравнениями. Точно так же z_{i1} и z_{i2} являются экзогенными переменными, потому что они определяются вне модели. ε_1 и ε_2 называются структурными ошибками.

Еще одно важное замечание: без z_1 или z_2 нет возможности определить, является ли уравнение уравнением спроса или предложения.

Рассмотрим следующую модель одновременных уравнений:

$$\begin{cases} y_1 = \alpha_1 y_2 + \beta_1 z_1 + \varepsilon_1 \\ y_2 = \alpha_2 y_1 + \beta_2 z_2 + \varepsilon_2 \end{cases}$$

В этой модели y_1 и y_2 являются эндогенными переменными, так как они определяются равновесием между двумя уравнениями; z_1 и z_2 являются экзогенными переменными. Поскольку z_1 и z_2 определяются вне модели, мы предполагаем, что z_1 и z_2 не коррелируют с обеими структурными ошибками.

Таким образом, по определению, экзогенные переменные в SEM также экзогенны в «эконометрическом смысле». Кроме того, две структурные ошибки ε_1 и ε_2 считаются некоррелированными друг с другом.

Теперь решим уравнения для y_1 и y_2 . Получим следующие уравнения приведенной формы:

$$\begin{cases} y_1 = \pi_{11} z_1 + \pi_{12} z_2 + v_1 \\ y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2 \end{cases},$$

где:

$$\pi_{11} = \frac{\beta_1}{1 - \alpha_1 \alpha_2},$$

$$\pi_{12} = \frac{\alpha_1 \beta_2}{1 - \alpha_1 \alpha_2},$$

$$v_1 = \frac{\varepsilon_1 + \alpha_1 \varepsilon_2}{1 - \alpha_1 \alpha_2},$$

$$\pi_{21} = \frac{\alpha_2 \beta_1}{1 - \alpha_2 \alpha_1},$$

$$\pi_{22} = \frac{\beta_2}{1 - \alpha_2 \alpha_1},$$

$$v_2 = \frac{a_2 \varepsilon_1 + \varepsilon_2}{1 - \alpha_2 \alpha_1}.$$

Эти параметры называются приведенными параметрами формы.

Можно проверить, что v_1 и v_2 не коррелируют с z_1 и z_2 . Поэтому можно оценить эти приведенные параметры формы с помощью МНК (просто примените МНК отдельно для каждого уравнения).

Однако нельзя оценить структурные уравнения по МНК. Рассмотрим первое структурное уравнение:

$$Y_1 = \alpha_1 y_2 + \beta_1 z_1 + \varepsilon_1.$$

Заметим, что

$$\text{cov}(y_2, \varepsilon_1) = \left[\frac{\alpha_2}{1 - \alpha_2 \alpha_1} \right] E(\varepsilon_1^2) = \left[\frac{\alpha_2}{1 - \alpha_2 \alpha_1} \right] \sigma_1^2 \neq 0.$$

Таким образом, y_2 коррелирует с ε_1 (предположим, что $\alpha_2 \neq 0$). Другими словами, y_2 является эндогенным в «эконометрическом смысле».

Таким образом, эндогенные переменные в SEM обычно эндогенны и в «эконометрическом смысле». То есть мы не можем применять OLS к структурным уравнениям.

Ковариацию между y_2 и ε_1 можно использовать для прогнозирования направления смещения:

$$\text{cov}(y_2, \varepsilon_1) = \left[\frac{\alpha_2}{1 - \alpha_2 \alpha_1} \right] \sigma_1^2.$$

Если она положительна, МНК-оценка α_1 будет смещена вверх. Если она отрицательна, будет смещена вниз. Вышеприведенная

формула не переносится на более общие модели. Но можно использовать ее как руководство для проверки направления смещения.

Ведущим методом оценки SEM является оценка IV.

Предположим, что мы имеем следующую модель, описывающую спрос и предложение:

$$\begin{cases} Q_t^S = \alpha_1 p + \beta_1 z_1 + \varepsilon \\ Q_t^D = \alpha_2 p + \varepsilon_1 \end{cases},$$

где первое уравнение — предложение, а второе уравнение — спрос.

Обратите внимание на то, что кривая предложения имеет наблюдаемый сдвиг предложения z_1 , но спрос не имеет наблюдаемого сдвига предложения.

Учитывая данные q , p и z_1 , мы должны решить, какое уравнение должно быть оценено. То есть мы должны решить, какое уравнение идентифицировано (рис. 19).

Поскольку есть наблюдаемый сдвиг предложения z_1 , который не содержится в уравнении спроса, можно идентифицировать уравнение спроса. В уравнении предложения имеется экзогенная переменная, позволяющая оценить уравнение спроса. В SEM иден-

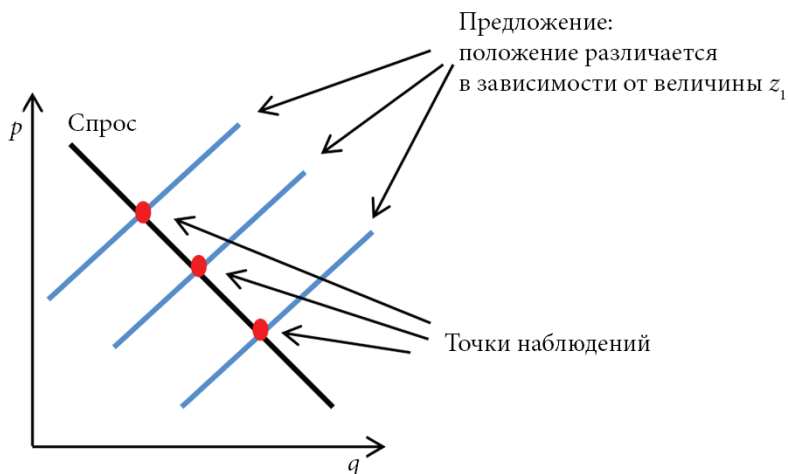


Рис. 19. Графическая иллюстрация модели

тификация используется для определения того, какое уравнение следует оценивать.

Теперь перейдем к более общему случаю. Предположим, что мы имеем следующую модель одновременного уравнения:

$$\begin{cases} y_1 = \beta_{10} + \alpha_1 y_2 + \beta_{11} z_{11} + \dots + \beta_{1k} z_{1k} + \varepsilon_1 \\ y_2 = \beta_{20} + \alpha_2 y_1 + \beta_{21} z_{21} + \dots + \beta_{2l} z_{2l} + \varepsilon_2 \end{cases},$$

где $(z_{11} \sim z_{1k})$ и $(z_{21} \sim z_{2l})$ могут содержать одни и те же переменные, но могут также содержать и разные переменные. Когда одно уравнение содержит экзогенные переменные, не содержащиеся в другом уравнении, это означает, что мы наложили ограничения на исключение.

Правило для идентификации: первое уравнение идентифицируется тогда и только тогда, когда второе уравнение содержит по крайней мере одну экзогенную переменную (ненулевой коэффициент), которая исключается из первого уравнения. Это условие имеет две компоненты. Во-первых, по крайней мере одна экзогенная переменная должна быть исключена из первого уравнения (условие порядка). Во-вторых, исключенная переменная должна иметь ненулевые коэффициенты во втором уравнении (условие ранга). Условие идентификации второго уравнения является просто зеркальным отображением указанного утверждения.

Пример: предложение рабочей силы замужних работающих женщин.

Уравнение предложения рабочей силы:

$$\begin{aligned} hours = & \alpha_1 lwage + \beta_{10} + \beta_{11} educ + \beta_{12} age + \beta_{13} kids + \\ & + \beta_{14} (NonWifeIncome) + \varepsilon_1. \end{aligned}$$

Уравнение предложения заработной платы:

$$lwage = \alpha_2 hours + \beta_{30} + \beta_{21} educ + \beta_{22} exp + \beta_{23} exp^2 + \varepsilon_2.$$

В модели часы (*hours*) и логарифм заработной платы (*lwage*) являются эндогенными переменными. Все остальные переменные яв-

ляются экзогенными. (Таким образом, мы игнорируем эндогенность образования (*educ*), возникающую из пропущенных переменных.)

Предположим, что нас интересует оценка первого уравнения. Поскольку *exr* и *exr*² исключены из первого уравнения, для первого уравнения выполняется условие порядка. Условие ранга состоит в том, что по крайней мере один из *exr* и *exr*² имеет во втором уравнении ненулевой коэффициент. Если предположить, что условие ранга выполнено, то первое уравнение будет идентифицировано. Аналогичным образом определяем, что второе уравнение также идентифицировано.

Как только мы определили, что уравнение идентифицировано, можем оценить его, используя двухшаговую процедуру наименьших квадратов.

Рассмотрим снова пример уравнения предложения рабочей силы. Нас интересует оценка первого уравнения. Предположим, что первое уравнение идентифицировано (выполняются условия порядка и ранга); *lwage* коррелирует с ϵ_1 . Таким образом, МНК использовать нельзя.

Тем не менее *exr* и *exr*² могут использоваться как инструменты для *lwage* в первом уравнении, так как *exr* и *exr*² не коррелируют с ϵ_1 по предположению модели (выполняется экзогенность инструмента). Также *exr* и *exr*² коррелируют с *lwage* по ранговому условию (релевантность инструмента выполняется). Следовательно, мы можем использовать исключенные экзогенные переменные в качестве инструментов.

Применение SEM к панельным данным. Рассмотрим следующую SEM:

$$\begin{cases} Y_{it1} = \alpha_1 y_{it2} + z_{it1} \beta_1 + (\alpha_{i1} + \epsilon_{it1}) \\ y_{it2} = \alpha_2 y_{it1} + z_{it2} \beta_2 + (\alpha_{i2} + \epsilon_{it2}) \end{cases}$$

Обозначение $z_{it1} \beta_1$ является сокращенным обозначением для

$$\beta_{11} z_{1t1} + \dots + \beta_{1k} z_{1tk} + \epsilon.$$

Обозначение $z_{it2}\beta_2$ является сокращенным обозначением для

$$\beta_{21}z_{2t1} + \dots + \beta_{2k}z_{2tk} + \varepsilon_2.$$

Из-за фиксированных эффектов α_{i1} и α_{i2} переменные коррелируют с составной ошибкой. Поэтому исключенные экзогенные переменные не могут использоваться в качестве инструментов. Чтобы применить 2SLS, мы должны использовать метод первых разностей или фиксированных эффектов.

Метод первых разностей:

$$\begin{cases} \Delta y_{it1} = \alpha_1 \Delta y_{it2} + \Delta z_{it1} \beta_1 + \Delta \varepsilon_{it1} \\ \Delta y_{it2} = \alpha_2 \Delta y_{it1} + \Delta z_{it2} \beta_2 + \Delta \varepsilon_{it2} \end{cases}$$

Метод фиксированных эффектов:

$$\ddot{y}_{it1} = \alpha_1 \ddot{y}_{it2} + \ddot{z}_{it1} \beta_1 + \ddot{\varepsilon}_{it1},$$

$$\ddot{y}_{it2} = \alpha_2 \ddot{y}_{it1} + \ddot{z}_{it2} \beta_2 + \ddot{\varepsilon}_{it2}.$$

Тогда Δz_{it1} , $\Delta z_{it2}\beta_2$ или \ddot{z}_{it1} , \ddot{y}_{it1} не коррелируют с ошибкой. Таким образом, мы можем применить метод 2SLS.

Процедура оценки такая же. Сначала определите, какое уравнение идентифицировано. Затем используйте исключенную экзогенную переменную в качестве инструментов в методе 2SLS.

Применение SEM к временным рядам. Одним из наиболее распространенных приложений 2SLS в прикладных исследованиях является оценка структурных моделей временных рядов.

Рассмотрим макромоделю совокупного спроса:

$$\begin{cases} C_t = \beta_0 + \beta_1 (Y_t - T_t) + \beta_2 r_t + \varepsilon_{1t} \\ I_t = \gamma_0 + \gamma_1 r_t + \varepsilon_{2t} \\ Y_t = C_t + I_t + G_t + \varepsilon_{3t} \end{cases}$$

В этой системе суммарный расход каждого квартала определяется совместно с располагаемым доходом. Если предположить, что налоги являются экзогенными (а на самом деле они восприим-

чивы к доходу), то функция потребления не может быть надежно оценена с помощью МНК. Если процентная ставка берется как экзогенная (например, благодаря денежно-кредитной политике), то инвестиции уравнения могут быть последовательно определены с помощью МНК. Третье уравнение представляет собой тождество; оно не обязательно должно быть оценено и выполняется без ошибок, но его наличие делает явным одновременный характер модели. Если r является экзогенной, то нам нужен один инструмент для оценки функции потребления; государственных расходов будет достаточно, и потребление будет точно определено. Если r следует рассматривать как эндогенную, то нужно добавить, по крайней мере, одно уравнение для модели, чтобы выразить, как денежно-кредитная политика реагирует на экономические условия. Мы могли бы также сделать инвестиционную функцию более реалистичной, добавив зависимость инвестиций от лаговых значений доходов, например, Y_{t-1} (фирмы составляют планы инвестиционных расходов, основанные на спросе на их продукцию). Это позволило бы использовать Y_{t-1} , предопределенную переменную, в качестве дополнительного инструмента для оценки функции потребления. Также можно использовать лаги экзогенных переменных, например, лаговые значения налогов или государственных расходов, как инструменты в данном контексте. Главное, что следует уяснить, — в этих моделях обычно требуются такие инструментальные переменные, как 2SLS, для последовательной оценки их компонентных соотношений.

ГЛАВА 4. ВРЕМЕННЫЕ РЯДЫ И ПАНЕЛЬНЫЕ ДАННЫЕ

Введение во временные ряды

Одной из целей анализа экономических данных является прогнозирование или предугадывание будущих значений экономических переменных. Один из способов сделать это — построить более или менее структурную эконометрическую модель, описывающую взаимосвязь между интересующей нас переменной и другими экономическими величинами для оценки этой модели по выборке данных, и использовать ее в качестве основы для прогнозирования и вывода. Хотя этот подход имеет преимущество, обеспечивая экономическую основу предсказаниям, он не всегда полезен. Например, возможно адекватное моделирование одновременных взаимосвязей между безработицей и инфляцией, но пока нельзя предсказать будущие темпы инфляции, также нельзя предсказать будущую безработицу.

Начнем с базовой модели:

$$y_t = \beta x_t + \varepsilon; \quad t=1, 2, \dots, T.$$

В целом размеры выборки, как правило, меньше в моделях временных рядов, чем в некоторых других рассмотренных нами моделях. Однако теперь нам нужно подумать о том, связаны ли наши ошибки последовательно. Ошибки последовательно коррелированы, если $E[\varepsilon_i, \varepsilon_j] \neq 0; i \neq j$. На практике мы рассматриваем только некоторые способы корреляции наших ошибок, в частности, пространственную автокорреляцию, корреляцию между пространственными единицами и последовательную корреляцию во времени. Здесь мы рассмотрим только вопросы временного ряда.

Временные ряды представляют собой последовательность данных, упорядоченных по времени. Когда мы говорим о переменных, которые могут меняться во времени, то должны упомянуть такие

экономические категории, как валовой внутренний продукт, инфляция, прибыль фирм и другие важные экономические условия.

Данные временных рядов подчиняются некоторым правилам и имеют особенности. Ключевой особенностью, усложняющей анализ, является тот факт, что экономические наблюдения редко (если вообще когда-либо) считаются независимыми во времени. Большинство экономических и других временных рядов связаны, часто сильно, с их недавним прошлым. Еще одной особенностью данных временных рядов, которая может потребовать особого внимания, является частота сбора данных. Время между y_t и y_{t+1} имеет определенную частоту. Она может быть разной: ежегодной, ежеквартальной, ежемесячной, ежедневной. Следующий момент — это временной интервал: период времени, в течение которого данные были собраны. Если в данных нет пробелов, мы можем найти временной интервал как число наблюдений, умноженное на частоту.

В общем случае временные ряды включают три элемента: детерминированный тренд, сезонную компоненту, циклическую компоненту. Тренд является основной особенностью временных рядов. Тренд может описываться различными уравнениями: линейными, логарифмическими, степенными и т. д. Фактический тип тренда устанавливается на основе его функциональной модели выборки статистическими методами или сглаживанием исходного времен-

ного ряда. Таким образом, можно описать тренд как $\sum_{i=0}^n a_i t^i$, если

$n = 1$, то имеем функцию линейного тренда, если $n = 2$, то существует квадратичная тенденция.

Сезонная компонента — это случай, когда некоторые переменные сильно зависят от определенного периода времени (сезонов). Например, это может быть число работников сельскохозяйственного сектора (потому что, если нет урожая, количество рабочих сокращается) или ежемесячное количество посетителей, приходящих на побережье, или количество потребителей сезонного продукта, например, мороженого.

Наконец, циклическая компонента является частью временного ряда, которая может быть описана ARMA (авторегрессион-

ное скользящее среднее или комбинация моделей авторегрессии и скользящего среднего) модели.

Простой способ моделирования зависимости между последовательными наблюдениями гласит, что Y_t линейно зависит от предыдущего значения Y_{t-1} . То есть

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t,$$

где ε_t обозначает серийно некоррелированные инновации со средним значением 0 и постоянной дисперсией. Процесс называется первым порядком процесса авторегрессии или AR(1) процессом. Из него следует, что текущее значение Y_t равно постоянной δ плюс θ раз своего предыдущего значения плюс непредсказуемая составляющая в данный момент ε_t . Мы рассматриваем процессы, которые и раньше обсуждали при автокорреляции (первого порядка) в модели линейной регрессии. Можно предположить, что $|\theta| < 1$. Процесс для ε_t является важным строительным блоком моделей временных рядов и называется процессом «белого шума». В этой главе ε_t всегда будет обозначать такой процесс, который гомоседастичен и не имеет автокорреляции. Теперь важно найти среднее, дисперсию и ковариацию временных рядов.

Ожидаемое значение Y_t можно найти из

$$E\{Y_t\} = \beta_0 + \beta_1 E\{Y_t - 1\}.$$

Очевидно, что $E\{Y_t\}$ не зависит от t , поэтому

$$\mu = E\{Y_t\} = \frac{\beta_0}{1 - \beta_1}.$$

Предположим, что $y_t = Y_t - \mu$, тогда наша модель временного ряда может быть переписана как

$$Y_t = \beta_1 y_{t-1} - 1 + \varepsilon_t,$$

$$E\{Y_t\} = E\{Y_{t-1}\}, \text{ if } |\beta_1| < 1.$$

Кроме того, можно определить

$$\text{cov}\{Y_t, Y_{t-1}\} = E\{y_t, y_{t-1}\} = E\{(\beta_1 y_{t-1} + \varepsilon_t) y_{t-1}\} = \beta_1 E\{y_{t-1}\} = \beta_1 \frac{\sigma^2}{1 - \beta_1^2}.$$

Обычно для всех k можно записать $\text{cov}\{Y_t, Y_{t-k}\} = \beta_1^k \frac{\sigma^2}{1 - \beta_1^2}.$

Пока θ отлична от нуля, любые два наблюдения на Y_t имеют ненулевую корреляцию; зависимость будет меньше (и потенциально произвольно близка к нулю), если наблюдения дальше друг от друга. Заметим, что ковариация между Y_t и Y_{t-k} зависит только от k , а не от t . Это отражает стационарность процесса.

Суммируем перечисленные особенности данных временных рядов:

- 1) частота: время между y_t и y_{t+1} : ежегодная, ежеквартальная, ежемесячная, ежедневная, ...;
- 2) временной интервал: период времени, в течение которого данные были собраны;
- 3) среднее $m_t = E(y_t)$;
- 4) дисперсия: $\text{var}(y_t) = E[(y_t - \mu_t)^2]$;
- 5) ковариация: $\text{cov}(y_t, y_{t-k}) = E[(y_t - \mu_t)(y_{t-k} - \mu_{t-k})]$.

Временные ряды могут быть стационарными или нестационарными. Если шоки происходят во времени t и со временем уменьшаются, а затем исчезают во времени $t + s$, где s стремится к ∞ , то такой временной ряд можно назвать стационарным. Также существуют некоторые формальные условия ковариационных стационарных временных рядов:

- 1) среднее значение для всех t и s с течением временем $\mu_t = \mu_{t-s} = \mu < \infty$;
- 2) дисперсия постоянна для всех t и s с течением временем $\text{var}(y_t) = \text{var}(y_{t-s}) = \sigma^2$;
- 3) ковариация во временных рядах не зависит от времени t , но на это влияет временной лаг (расстояние между двумя элементами временных рядов):

$$\text{cov}(y_t, y_{t-s}) = \text{cov}(y_{t-j}, y_{t-j-s}) = \gamma_s.$$

Когда мы тестируем последовательную корреляцию, мы обычно предполагаем что-то о процессе ошибки. Обычными предположениями являются ошибки, связанные либо с процессом авторегрессии первого порядка AR(1), либо с процессом скользящего среднего первого порядка MA(1).

AR(1) процесс: $\varepsilon_t = p\varepsilon_{t-1} + v_t$, текущая ошибка является частью предыдущей ошибки плюс некоторый шок. Это процесс первого порядка, потому что ε_t зависит только от его ближайшего лагового значения. Очевидно, что это модель экспоненциального затухания. Например:

$$\varepsilon_t = p\varepsilon_{t-1} + v_t,$$

$$\varepsilon_t = p(p\varepsilon_{t-2} + v_{t-1}) + v_t,$$

$$\varepsilon_t = p^2(p\varepsilon_{t-3} + v_{t-2}) + pv_{t-1} + v_t,$$

$$\varepsilon_t = p^3(p\varepsilon_{t-3} + p^2v_{t-2}) + pv_{t-1} + v_t \dots$$

Текущая ошибка — это всего лишь ошибка всех предыдущих запасов, взвешенная некоторым коэффициентом (p), экспоненциально снижающимся. Как быстро эффект этих предыдущих ошибок закончится, зависит от значения p .

Когда $p < 1$, то мы имеем так называемые стационарные временные ряды. В этой установке временные ряды выглядят зубчатыми, у нее есть структура, и она никогда не блуждает слишком далеко от среднего. Эффект ошибок затухает и исчезает (не полностью) со временем. Вещи, которые происходят в последнее время, относительно важнее, чем события, которые произошли давно.

Когда $p > 1$, то мы имеем так называемые нестационарные временные ряды. В данном случае получаем более плавный временной ряд, который в конце концов взрывается. Вещи, которые произошли давно, имеют большее влияние по сравнению с тем, что произошло совсем недавно.

Когда $p = 1$, то мы имеем «случайное блуждание», т. е. $y_t = y_{t-1} + \varepsilon_t$. В данном случае временные ряды перемещаются вверх и вниз,

но медленно. То, что произошло в любой период, имеет одинаковое влияние.

МА(1) процесс: $\varepsilon_t = \mu v_{t-1} + v_t$, где $v_t \sim iid$. Таким образом, текущая ошибка действительно является частью ошибки за предыдущий период плюс ошибка этого периода. Процесс МА — это просто линейная комбинация «белых шумов» или *iid*-ошибок. В отличие от процесса ошибки AR(1) эффект процесса ошибки МА(1) кратковременный и конечный — на него влияют только текущие и предыдущие значения v .

Можно заключить, что временные ряды могут быть стационарными или нестационарными. Когда мы исследуем нестационарные временные ряды, нужно помнить о проблеме единичного корня и о том, как определить существование этой проблемы. Единичный корень — это термин, используемый при анализе временных рядов, характеризующий свойство некоторых нестационарных временных рядов. Название связано с тем, что так называемое характеристическое уравнение (или характеристический многочлен) авторегрессионной модели временных рядов имеет корни, равные единице. Наличие единичных корней в авторегрессии модели временных рядов эквивалентно понятию интегрирования временного ряда.

Интегрированный временной ряд — это нестационарный временной ряд, разности некоторого порядка от которого являются стационарным временным рядом. Такие ряды также называют *разностно-стационарными* (*DS-рядами*, *Difference Stationary*). Примером интегрированного временного ряда является случайное блуждание, часто используемое при моделировании финансовых временных рядов.

Понятие интегрированных временных рядов тесно связано с единичными корнями в моделях авторегрессии. Наличие единичных корней в характеристическом уравнении авторегрессионной составляющей модели временных рядов относится к интеграции временных рядов. И число единичных корней имеет тот же порядок интегрирования.

Важным понятием, связанным с временными рядами, является понятие единичного корня. Предположим, что мы имеем линейную функцию модели временного ряда с простым AR(1) процессом, где

ε_t — независимые нормальные ошибки с нулевым средним и дисперсией σ^2 : $\{y_t\}_{t=1}^T$, $y_t = a_1 y_{t-1} + \varepsilon_t$.

Первые разностные временные ряды являются стационарными. Представим модель в несколько иной форме:

$$\begin{aligned} y_t &= a_1 y_{t-1} + \varepsilon_t = a_1 + a_1 + y_{t-2} + \varepsilon_t + \varepsilon_{t-1} = \\ &= a_1 + a_1 + a_1 + y_{t-3} + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} = y_0 + at + \sum_{i=1}^t \varepsilon_i. \end{aligned}$$

Таким образом, случайное блуждание с тенденцией внешне похоже на модель линейного тренда с одним очень важным отличием

чем $V\left(\sum_{i=1}^t \varepsilon_i\right) = t\sigma^2$, дисперсия ошибки модели пропорциональна

времени, т. е. со временем она стремится к бесконечности. Хотя математическое ожидание случайной ошибки равно нулю. Даже применительно к временному ряду в процессе устранения линейного (детерминированного) тренда мы все равно получаем нестационарный процесс, стохастический тренд.

В данном случае временные ряды будут стационарными, если $|a_1| < 1$, и содержать единичный корень при $|a_1| = 1$. Но ряд, содержащий единичный корень, не является стационарным. Таким образом, мы можем видеть, что оценки, генерируемые методом МНК, являются предвзятыми.

В более сложном случае нужно использовать специальный тест для обнаружения единичного корня во временных рядах. Есть много тестов: тест Филлипса — Перрона, тест Лейбурна, тест Шмидта — Филлипса, тест Кохрейна и др. Наиболее широко используемым является тест Дики — Фуллера и расширенный тест Дики — Фуллера. Чтобы показать, как это работает, используем данный авторегрессионный процесс AR(1): $y_t = a_1 y_{t-1} + \varepsilon_t$.

Если вычесть y_{t-1} из обеих сторон этого уравнения, получим $\Delta y_t = \gamma y_{t-1} + \varepsilon_t$, где $\gamma = a_1 - 1$. Поэтому проверка гипотезы единичного корня в этом случае подразумевает проверку нулевой гипотезы равенства нулю нуля коэффициента γ . В случае ликвидации «взрыв-

ного» процесса тест является односторонним, т. е. альтернативная гипотеза — это гипотеза о том, что коэффициент γ меньше нуля.

$$H_0: a_1 = 1 \text{ или } \gamma = 0,$$

$$H_a: a_1 < 1 \text{ или } \gamma < 0.$$

Тестовая статистика (*DF-статистика*) является обычной t -статистикой для проверки значимости коэффициентов линейной регрессии. Однако распределение этой статистики отличается от классического распределения t -статистики (распределения Стьюдента или асимптотического нормального распределения). Распределение *DF*-статистики выражается через винеровский процесс и называется распределением Дики — Фуллера.

Существуют три версии теста (тестовые регрессии):

1) нет константы, нет трендов:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t;$$

2) существует константа, но тренд отсутствует:

$$\Delta y_t = \mu + \gamma y_{t-1} + \varepsilon_t;$$

3) существуют константа и тренд:

$$\Delta y_t = \mu + \beta t + \gamma y_{t-1} + \varepsilon_t.$$

Для каждой из трех тестовых регрессий существуют критические значения *DF*-статистики, которые берутся из специальных таблиц Дики — Фуллера. Если значение статистики лежит слева от критического значения (критическое значение отрицательно) на заданном уровне значимости, нулевая гипотеза единичного корня отвергается, и процесс допускает стационарность (в контексте данного теста). В противном случае гипотеза не будет отвергнута, и процесс может содержать единичные корни, т. е. быть нестационарным (интегрированным) временным рядом (табл. 5).

Формализуем:

$$H_0: \gamma = 0 \text{ (присутствует единичный корень);}$$

$$H_a: \gamma < 0 \text{ (стационарный временной ряд).}$$

$$t_{DF} = \frac{\hat{\gamma}}{s.e.\hat{\gamma}}.$$

Правило отклонения: $t_{DF} < t_{cr}$.

Таблица 5

**Критическое значение статистики
теста Дики — Фуллера на уровне значимости 1 %**

Sample	Model № 1	Model № 2	Model № 3
25	-2,66	-3,75	-4,38
50	-2,62	-3,58	-4,15
100	-2,60	-3,51	-4,04
∞	-2,58	-3,43	-3,96

Если в регрессионном тесте добавлены лаги первых разностей временного ряда, распределение DF -статистики (и, следовательно, критического значения) не изменится. Этот тест называется *расширенным тестом Дики — Фуллера (Augmented DF или ADF)*.

Необходимость включения лагов первых разностей заключается в том, что процесс может быть авторегрессией первого и более высокого порядков. Механизм тот же, но модели разные. Включая лаги разностей в оценочное уравнение, можно исправить возможную автокорреляцию в остатках.

Существуют три версии теста (тестовые регрессии):

1) нет константы и трендов:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^k \rho_i \Delta y_{t-i} + \varepsilon_t;$$

2) есть константа, тренд отсутствует:

$$\Delta y_t = \mu + \gamma y_{t-1} + \sum_{i=1}^k \rho_i \Delta y_{t-i} + \varepsilon_t;$$

3) константа и тренд существуют:

$$\Delta y_t = \mu + \beta t + \gamma y_{t-1} + \sum_{i=1}^k \rho_i \Delta y_{t-i} + \varepsilon_t.$$

Чтобы узнать, сколько лагов следует включить, начните с $T/4$. Проверьте значение наивысшего лага и уменьшите его, если оно незначимо. Помните, что чем выше частота, тем больше K .

Тест Дики — Фуллера, как и многие другие тесты, проверяет наличие только одного единичного корня. Однако теоретически процесс может иметь несколько единичных корней. В этом случае тест может быть неправильным.

Тест Дики — Фуллера, или *DF*-тест, используется для любой из процедур оценки, описанных выше, и поэтому может основываться на регрессии с трендом или без него. Если графический анализ ряда указывает на четкую положительную или отрицательную тенденцию, это наиболее подходящий вариант для проведения теста Дики — Фуллера с трендом. Это означает, что альтернативная гипотеза позволяет процессу демонстрировать линейный детерминированный тренд. Важно подчеркнуть, что гипотеза единичного корня соответствует нулевой гипотезе. Если мы не можем отказаться от присутствия единичного корня, это не обязательно означает, что гипотеза истинна. Такой факт можно объяснить недостаточностью информации в данных, чтобы отклонить ее. Это общая разница между принятием гипотезы и отказом от нее. Поскольку долгосрочные свойства процесса в решающей степени зависят от наложения единичного корня, об этом нужно знать.

Не все ряды, для которых мы не можем отказаться от гипотезы единичного корня, обязательно интегрированы первым порядком. Чтобы обойти проблему, связанную с тем, что тесты с единичным корнем часто имеют низкую мощность, Квятковский, Филлипс, Шмидт и Шин (1992) предложили альтернативный тест, где стационарность является нулевой гипотезой, а существование единичного корня является альтернативной. Данный тест обычно называют *KPSS-тестом*. Основная идея заключается в том, что временной ряд разлагается на сумму детерминированного тренда времени, случайного блуждания и стационарной ошибки (обычно это не «белый шум»). Нулевая гипотеза (стационарности тренда) указывает, что дисперсия компоненты случайного блуждания равна нулю. Тест на самом деле является критерием множителя Лагранжа, и вычислить тестовую статистику довольно просто. Во-первых, исследователь должен провести вспомогательную регрессию Y_t на точку пересечения и временной тренд t . Затем сохранить остатки по МНК

и вычислить частичные суммы $S_t = \sum_{s=1}^t e_s$ для всех t . Тогда тестовая статистика представляется как

$$KPSS = \sum_{t=1}^T \frac{S_t^2}{\hat{\sigma}^2},$$

где $\hat{\sigma}^2$ является оценкой дисперсии ошибок. Эта последняя оценка $\hat{\sigma}^2$ может включать поправки для автокорреляции, основанные на формуле Ньюи–Веста. Асимптотическое распределение является нестандартным, и Квятковский сообщает о 5 %-м критическом значении 0,146. Если нулевая гипотеза — это стационарность, а не стационарность тренда, то тренд следует исключить из вспомогательной регрессии. Затем статистическую статистику вычисляют таким же образом, но 5 %-е критическое значение составляет 0,463.

AR-, MA-, ARMA-модели

Шестое условие Гаусса — Маркова часто нарушается в данных временных рядов. В этом случае ошибки для разных наблюдений коррелируют друг с другом: $\text{corr}(\varepsilon_i, \varepsilon_j) \neq 0$, если $i \neq j$. В эконометрике это называется последовательной корреляцией или автокорреляцией. Существует шаблон по всем ошибкам. Тогда ошибки не распределяются независимо друг от друга по наблюдениям и не являются строго случайными. Основные свойства автокорреляции описаны в главе 3 данной работы.

Стационарный процесс авторегрессии (AR). Многие наблюдаемые временные ряды демонстрируют серийную автокорреляцию, т. е. линейную связь между лаговыми наблюдениями. Это означает, что прошлые наблюдения могут предсказать текущие наблюдения. Процесс авторегрессии (AR) моделирует условное среднее y_t как функцию прошлых наблюдений:

$$y_{t-1}, y_{t-2}, \dots, y_{t-p}.$$

Процесс AR, который зависит от p предыдущих наблюдений, называется AR-моделью порядка p , обозначаемой AR(p).

Стационарные авторегрессивные процессы (AR) имеют теоретические автокорреляционные функции (ACF), которые распадаются до нуля, вместо того, чтобы отсекают нуль. Коэффициенты автокорреляции могут часто чередоваться в знаке или показывать волнообразный рисунок, но во всех случаях они уходят в нуль. Напротив, процессы AR с порядком p имеют теоретические частные автокорреляционные функции (PACF), которые обрезаются до нуля после запаздывания p . Длина лага конечного пика PACF равна AR-порядку процесса (p).

Авторегрессионный процесс порядка p , AR(p):

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \varepsilon_t.$$

Процесс скользящего среднего (МА). Модель скользящего среднего (МА) фиксирует последовательную автокорреляцию во временном ряду y_t , выражая условное среднее y_t как функцию прошлых нововведений:

$$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}.$$

Модель МА, которая зависит от q прошлых нововведений, называется МА-моделью порядка q и обозначается как МА(q).

Теоретические функции ACF-процессов МА (скользящего среднего) с порядком q обрезаются до нуля после лага q . Однако их теоретические PACF распадаются до нуля. Длина лага конечного пика ACF равна порядку МА процесса (q).

Процесс скользящего среднего порядка q , МА(q): $y_t = \sum_{i=1}^q \beta_i \varepsilon_{t-i}$, β_0 , обычно нормируется к 1.

Стационарный смешанный процесс (ARMA). Для некоторых наблюдаемых временных рядов необходима модель более высокого порядка AR или МА, чтобы смоделировать базовый процесс. В этом случае комбинированная модель авторегрессии — скользящего среднего (ARMA) иногда может быть более экономным выбором.

Модель ARMA выражает условное среднее y_t как функцию и прошлых наблюдений $(y_{t-1}, \dots, y_{t-p})$, и прошлых нововведений: $(\varepsilon_{t-1}, \dots, \varepsilon_{t-q})$.

Количество прошлых наблюдений, зависящих от y_t , p , является порядком AR; количество прошлых нововведений, зависящих от y_t , q , — порядком MA. В общем случае эти модели обозначаются как ARMA(p, q).

Процесс авторегрессии — скользящего среднего порядков p и q , ARMA(p, q):

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i}.$$

Для оценки модели ARMA ряды должны быть стационарными. Стационарность временных рядов тесно связана с понятием устойчивости однородных линейных разностных уравнений.

Если линейное однородное разностное уравнение устойчиво, то его решение сходится к нулю при t , стремящемся к бесконечности. Если оно нестабильно, то его решение расходится.

Условия устойчивости для процесса AR(p)

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \varepsilon_t.$$

Приведем линейное однородное разностное уравнение p -го порядка:

$$\alpha^p - a_1 \alpha^{p-1} - a_2 \alpha^{p-2} - \dots - a_p = 0.$$

Характеристические корни данного уравнения: $\{\alpha^i\}_{i=1}^p$.

Стабильность соответствующего p -го порядка линейного однородного разностного уравнения является необходимым и достаточным условием стационарности созданных временных рядов:

$$\sum_{i=1}^p \alpha_i < 1 \text{ является необходимым условием устойчивости;}$$

$\sum_{i=1}^p |\alpha_i| < 1$ является достаточным условием устойчивости;

$\sum_{i=1}^p \alpha_i = 1$ — разностное уравнение содержит единичный корень,

это означает, что хотя бы один из характеристических корней равен единице.

Условия устойчивости для процесса $MA(q)$

$$y_t = \sum_{i=1}^q \beta_i \varepsilon_{t-i}.$$

Приведем однородное разностное уравнение, которое устойчиво:

$$y_i = 0.$$

— Если q конечно, то созданный временной ряд будет стационарным.

— Если q бесконечно, то необходимым и достаточным условием стационарности созданного временного ряда является то, что суммы

$$\beta_s + \beta_1 \beta_{s+1} + \beta_2 \beta_{s+2} + \dots$$

конечны для всех s .

Условия устойчивости для процесса $ARMA(p, q)$

$$y_t = \alpha_0 + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i}.$$

Необходимыми и достаточными условиями стационарности созданных временных рядов являются следующие:

— линейное однородное разностное уравнение соответствующего p -го порядка устойчиво;

— суммы $\beta_s + \beta_1 \beta_{s+1} + \beta_2 \beta_{s+2} + \dots$ конечны для всех s .

Стационарность временного ряда требует, чтобы его среднее значение, дисперсия и ковариации были постоянными.

Автокорреляция и частичная автокорреляция. Функции ACF и PACF

Функции ACF и PACF помогают идентифицировать параметры p и q модели ARMA(p, q).

Автокорреляция — это линейная зависимость переменной от самой себя в два момента времени. Для стационарных процессов автокорреляция между любыми двумя наблюдениями зависит только от временного лага h между ними. Обозначим

$$\text{cov}(y_t, y_{t-h}) = \gamma_h.$$

Автокорреляция лага h определяется как

$$\rho_h = \text{corr}(y_t, y_{t-h}) = \frac{\gamma_h}{\gamma_0}.$$

Знаменатель γ_0 является ковариацией лага 0, другими словами, это безусловная дисперсия процесса.

Корреляция между двумя переменными может быть результатом взаимной линейной зависимости от других переменных. Частичная автокорреляция представляет собой автокорреляцию между y_t и y_{t-h} после удаления любой линейной зависимости из $(y_1, y_2, \dots, y_{t-h+1})$. Частичная автокорреляция лага h обозначается через $\phi_{h,h}$.

Автокорреляционная функция (ACF) для временного ряда $y_t (t=1, \dots, N)$ является последовательностью $\rho_h (h=1, 2, \dots, N-1)$. Частичная автокорреляционная функция (PACF) является последовательностью $\phi_{h,h} (h=1, 2, \dots, N-1)$.

Теоретические ACF и PACF для условных средних моделей AR, MA и ARMA conditional известны и весьма различны для каждой модели. Различия в ACF и PACF полезны при выборе моделей. Ниже приводится описание поведения ACF и PACF для этих моделей.

Метод Бокса — Дженкинса, названный в честь Джорджа Бокса и Гвилима Дженкинса, заключается в применении авторегрессивных скользящих средних ARMA или ARIMA (интегрированные модели порядка n), чтобы найти наилучшее соответствие модели временного ряда прошлым значениям временного ряда. Другими

словами, методология Бокса — Дженкинса используется для оценки соответствующей модели $ARMA(p, q)$.

Методология Бокса — Дженкинса представляет собой четырехэтапный процесс идентификации, выбора и оценки условных средних моделей (для дискретных, одномерных данных временных рядов).

1. *Установление стационарности временного ряда.* Если ряд нестационарный, для достижения стационарности исследователь должен успешно выполнять разность рядов. Автокорреляционная функция (ACF) и частичная автокорреляционная функция (PACF) выборки стационарного ряда затухают экспоненциально (или полностью прерываются после нескольких лагов).

2. *Идентификация.* Нужно построить графики ACF и PACF для лагов до $T/4$, после этого выбрать соответствующее количество лагов p и q .

3. *Оценка.* На этом шаге оценивается выбранная $ARMA(p, q)$, проверяются условия устойчивости и сохраняются остатки.

4. *Диагностическая проверка.* Строятся графики ACF и PACF для серии остатков, после чего вычисляется Q-статистика и, наконец, выполняются Q-тесты.

— Если все автокорреляции выборки и частичные автокорреляции близки к нулю и если все Q-тесты не отклонят нулевую гипотезу об отсутствии автокорреляции, то оценочная модель может быть правильной.

— Если нет, вернитесь к шагу 1 и измените количество лагов p и q .

Оценка параметров моделей Бокса — Дженкинса является довольно сложной задачей нелинейного оценивания. По этой причине оценка параметров должна быть предоставлена программному обеспечению высокого качества, которое подходит для моделей Бокса — Дженкинса. К счастью, многие статистические программы теперь подходят для моделей Бокса — Дженкинса.

Основными подходами для изучения моделей Бокса — Дженкинса являются нелинейные наименьшие квадраты и оценка максимального правдоподобия. Оценка максимального правдоподобия обычно является предпочтительным методом. Уравнения правдоподобия

для полной модели Бокса — Дженкинса сложны и не включены в главу.

Введение в панельные данные

Панельные или пространственно-временные данные представляют собой наблюдение за некоторыми переменными на разных объектах (например, разных фирмах или регионах) во времени. Формально мы можем написать это как

$$x_{it} = \begin{pmatrix} 1 \\ x_{lit} \\ x_{kit} \end{pmatrix},$$

где $i = 1, \dots, n$ и $t = 1, \dots, t$.

Данные панели позволяют вам контролировать переменные, которые вы не можете наблюдать или измерять (например, культурные факторы или различия в деловой практике между компаниями); или переменные, которые изменяются с течением времени, но не по всем объектам (например, национальная политика, федеральные правила, международные соглашения и т. д.). Это объясняет индивидуальную гетерогенность. С панельными данными вы можете включать переменные на разных уровнях анализа (например, студенты, школы, районы, штаты), подходящие для многоуровневого или иерархического моделирования. Некоторыми недостатками являются проблемы сбора данных (например, проектирование выборки, охват), отсутствие ответа в случае микропанелей или пространственной зависимости в случае макропанелей (т. е. корреляции между странами).

Оценка панельных данных имеет некоторые важные преимущества:

1) проведение исследования экономических вопросов, которые могут быть проанализированы с помощью сочетания пространственных данных и временных рядов (например, анализ прибыли фирм за весь производственный период);

2) увеличение количества точек данных, более точные результаты наблюдений;

3) возможность идентификации гетерогенности в интересующих объектах;

4) увеличение степеней свободы. Коллинеарность среди объясняющих переменных становится ниже, а эффективность оценок улучшается;

5) доступность нескольких наблюдений для данного индивида или в заданное время, позволяющая идентифицировать неидентифицируемую модель;

6) облегчение контроля последствий пропущенных или ненаблюдаемых переменных.

Проблемы с панельными данными делятся на три группы:

1) проблемы проектирования и сбора данных:

— проблемы охвата (неполный учет населения, представляющего интерес),

— отсутствия ответа (из-за отсутствия сотрудничества респондента или из-за ошибки интервьюера),

— отсутствия информации (респондент не запоминает данные или не помнит неправильно),

— частоты опроса;

2) проблемы искажения ошибок измерения. Ошибки измерения могут возникать из-за ошибочных ответов, вызванных:

— непонятными вопросами,

— ошибками памяти,

— неподходящими информаторами,

— неверной записью ответов,

— эффектами интервьюера;

3) проблемы селективности:

— самоотбор. В статистике смещение из-за самоотбора возникает в любой ситуации, когда индивиды выбирают себя в группу, что приводит к необъективной выборке без вероятности. Оно обычно используется для описания ситуаций, когда характеристики людей, которые заставляют их выбирать себя в группу, создают ненормальные или нежелательные условия в группе,

— отсутствие ответа (отказались отвечать, никого нет дома ...),

— респонденты могут умереть, или переехать, или решить, что цена ответа высока для них.

Одной из наиболее важных проблем, связанных с использованием данных панели, является селективное смещение. Это ошибка, возникающая в случае отбора наблюдений, проблема смещенного (неслучайного) выбора данных. Неслучайный отбор — это основание для смещения МНК-оценок.

Одной из наиболее важных проблем панельных данных является селективное смещение. Это означает, что мы отбираем выборку не случайным образом, а предвзято. Это может быть причиной смещения оценок наименьших квадратов. Предположим, что у нас есть модель регрессии, в которой y — уровень образования:

$$y_i = \beta'x_i + u_i, i = 1, \dots, N.$$

Иногда по разным причинам нас интересует анализ выборки с определенным уровнем образования, поэтому

$$y_i = \beta'x_i + u_i > L, \text{ включен,}$$

$$y_i = \beta'x_i + u_i < L, \text{ исключен.}$$

В этом случае можно получить предвзятую оценку. Выборка с уровнем образования ниже требуемого уровня L будет исключена из данного исследования. Оценка по МНК будет недооценивать эффекты. Другими словами, процедура отбора выборки вводит корреляцию между переменными в правой части и ошибкой, что приводит к смещению вниз линии регрессии.

Говоря об ограничениях панельных данных, мы также должны сказать несколько слов о гетерогенном смещении. Игнорирование индивидуальных или зависящих от времени эффектов, которые существуют в единицах пространственного или временного ряда, но не учитываются включенными пояснительными переменными, может привести к параметрической гетерогенности в спецификации модели.

Игнорирование такой гетерогенности (в наклоне и/или константе) может привести к несостоятельным или бессмысленным оценкам анализируемых параметров.

Предположим, что мы имеем регрессионную модель $y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}$, тогда модель $y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}$ должна быть гомогенной моделью.

Очевидно, что в данных случаях объединенная регрессия, игнорирующая гетерогенные точки пересечения, никогда не должна использоваться. Более того, направление смещения оценок объединенного наклона не может быть идентифицировано априорно.

Чтобы исправить некоторую предвзятость, можно использовать объединенную модель с объединенным МНК. Наиболее ограниченной моделью, в которой указаны сокращенные коэффициенты, является следующая: $y_{it} = \alpha + x'_{it}\beta + z'_i\gamma + c_i + u_{it}$.

Если модель указана правильно и регрессоры не связаны с ошибкой, то ее можно оценить с использованием объединенного МНК. В этом случае в модели y_{it} — зависимая переменная, x'_{it} — это K -пространственный вектор-строка изменяющихся во времени объясняющих переменных и z'_i — M -мерный вектор строки нестационарных объясняющих переменных, исключающих константу, α — точка пересечения, β — K -мерный вектор-столбец параметров, γ — M -мерный вектор-столбец параметров, c_i является индивидуально-специфическим эффектом и u_{it} — идиосинкразическая ошибка.

NT наблюдений для всех индивидов и периодов времени:

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_{NT} \end{bmatrix}_{NT \times 1} \quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_{NT} \end{bmatrix}_{NT \times 1} \quad Z = \begin{bmatrix} z_1 \\ \vdots \\ z_i \\ \vdots \\ z_{NT} \end{bmatrix}_{NT \times 1} \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_i \\ \vdots \\ u_{NT} \end{bmatrix}_{NT \times 1}$$

Объединенная оценка по МНК игнорирует структуру панельных данных и просто оценивает α , β и γ как

$$\begin{pmatrix} \hat{\alpha}_{POLS} \\ \hat{\beta}_{POLS} \\ \hat{\gamma}_{POLS} \end{pmatrix} = (W'W)^{-1} W'y,$$

где $W = [I_{NT} XZ]$, I_{NT} — это вектор $NT \times 1$ единиц.

Вероятно, ошибка будет коррелирована с течением времени для данного человека, однако в этом случае обычные стандартные ошибки не должны использоваться, поскольку они могут быть сильно смещены вниз. Объединенная МНК-оценка несостоятельна, если есть подходящая модель фиксированных эффектов (см. следующую главу).

Виды оценок, применяемых для панельных данных

Предположим, что у нас есть данные по каждой пространственной единице за T -периоды времени. Это *модель ненаблюдаемых эффектов (UEM)*, также называемая моделью ошибочных компонент. Мы можем записать модель для каждого периода времени:

$$y_{i1} = x_{i1}\beta + c_i + \varepsilon_{i1},$$

$$y_{i2} = x_{i2}\beta + c_i + \varepsilon_{i2},$$

...

$$y_{iT} = x_{iT}\beta + c_i + \varepsilon_{iT},$$

где t — наблюдения переменной y для индивида i ; x_{it} — вектор объясняющих переменных, измеренный во времени t ; c_i — не наблюдается во всех периодах, но постоянна во времени; ε_{it} — является изменяющейся во времени идиосинкразированной ошибкой.

Оценка первых разностей (FD). Уравнение первых разностей, которое мы получаем дальше, очень просто. Это всего лишь одно пространственное уравнение, но каждая переменная отличается

во времени. Мы можем проанализировать его с использованием методов, которые разработали ранее, при условии соблюдения ключевых допущений. Наиболее важным из них является то, что Δu_i не коррелирует с Δx_i . Это предположение справедливо, если идиосинкратическая ошибка в каждый момент времени t , u_{it} , не коррелирует с объясняющей переменной в каждый период времени. Это еще одна версия предположения о строгой экзогенности для моделей временных рядов. В частности, это предположение исключает случай, когда x_{it} является лаговой зависимой переменной. Хотя возможен вариант, чтобы x_{it} коррелировала с ненаблюдаемыми факторами, которые являются постоянными во времени. Полученную МНК-оценку мы называем *оценкой первых разностей*.

Предположим, что есть модель, и она имеет следующее значение в предыдущем периоде:

$$y_{it} = x_{it}\beta + c_i + \varepsilon_{it}, t = 1, \dots, T,$$

$$y_{it-1} = x_{it-1}\beta + c_i + \varepsilon_{it}, t = 2, \dots, T.$$

Примечание. Любые постоянные, в течение всего времени объясняющие переменные, удаляются вместе с c_i . Следовательно, вам нужно выбрать объясняющие переменные, которые меняются со временем.

Вычитая второе уравнение из первого, имеем:

$$\Delta y_{it} = y_{it} - y_{it-1} = \Delta x_{it}\beta + \Delta \varepsilon_{it}, t = 2, \dots, T.$$

После преобразования ненаблюдаемые c_i исключаются из модели.

FD -оценка $\hat{\beta}_{FD}$ получается путем регрессирования изменений в изменениях с использованием МНК:

$$\hat{\beta}_{FD} = (\Delta X^T \Delta X)^{-1} \Delta X^T \Delta y.$$

Заметим, что условие ранга должно выполняться для $\Delta X^T \Delta X$, чтобы была обратимость: $\text{rank}(\Delta X^T \Delta X) = k$.

По аналогии

$$\widehat{\text{Avar}}(\hat{\beta}_{FD}) = \hat{\sigma}_{\varepsilon}^2 (\Delta X^T \Delta X)^{-1},$$

где $\hat{\sigma}_{\varepsilon}^2$ получается из

$$\hat{\sigma}_{\varepsilon}^2 = [n(T-1) - K]^{-1} \hat{\varepsilon}^T \hat{\varepsilon}.$$

Предположения для объединенного МНК с использованием первых разностей. Для указанной выше модели и случайной пространственной выборки должны быть выполнены следующие допущения:

— Каждая объясняющая переменная изменяется со временем (по крайней мере, для некоторого i), и нет идеальных линейных отношений между объясняющими переменными.

— Для каждого t ожидаемое значение идиосинкразической ошибки с учетом объясняющих переменных во все периоды времени и ненаблюдаемого эффекта равно нулю: $E(u_{it} | x_i, a_i) = 0$. В этом случае мы иногда говорим, что x_{ij} строго экзогенны для ненаблюдаемого эффекта. Идея состоит в том, что если мы контролируем a_i , то нет никакой корреляции между x_{ij} и оставшейся идиосинкразированной ошибкой, u_{it} , для всех s и t .

— Дисперсия разностных ошибок, обусловленных всеми объясняющими переменными, является постоянной:

$$\text{var}(\Delta u_{it} | x_i) = \sigma^2, t = 2, \dots, T.$$

— Для всех $t \neq s$ разности в идиосинкразических ошибках некоррелированы (обусловлены всеми объясняющими переменными):

$$\text{cov}(\Delta u_{it}, \Delta u_{is} | x_i) = 0, t \neq s.$$

Условно x_i и Δu_{it} являются независимыми и одинаково распределенными нормальными случайными величинами.

Оценка фиксированных эффектов (FE). Когда существует ненаблюдаемый фиксированный эффект, альтернативой первым разностям является оценка фиксированных эффектов. Методы фиксированных эффектов также преобразуют модель с удалением

c_i перед оценкой. Исключение c_i необходимо, поскольку она может быть коррелирована x_{it} , следовательно:

$$\text{cov}(c_i, x_{it}) \neq 0.$$

Таким образом, в модели существует проблема эндогенности.

Давайте вернемся к модели, представленной в начале данной главы. Среднее уравнение

$$y_{it} = x_{it}\beta + c_i + \varepsilon_{it}$$

после преобразования со временем выглядит так:

$$\bar{y}_i = \beta \bar{x}_i + c_i + \bar{\varepsilon}_i.$$

Заметим, что

$$\bar{y}_i = \frac{\sum_{t=1}^T y_{it}}{T}.$$

Вычтем второе уравнение из первого, получим:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + \varepsilon_{it} - \bar{\varepsilon}_i.$$

Объединенная МНК-оценка, основанная на понижении раз-
мерности переменных, называется оценкой с фиксированными
эффектами или внутригрупповой — within estimator (МНК ис-
пользует изменение во времени y и x в каждом пространственном
наблюдении).

Предположим, что у нас есть регрессия фиктивной переменной:

$$y_{it} = x_{it}\beta + c_i + \varepsilon_{it}.$$

Традиционный взгляд FE: предполагается, что ненаблюдаемый
эффект c_i можно оценить для каждого i (т.е. c_i — индивидуальная
точка пересечения).

П р и м е ч а н и е. Мы не можем выполнить это для одного про-
странства — будет $(N + k)$ параметров для оценки и только N на-
блюдений.

Регрессия фиктивной переменной — это способ оценки точки пересечения для каждого i путем включения фиктивной переменной для каждого пространственного наблюдения вместе с объясняющими переменными.

Первые разности vs фиксированные эффекты. Первые разности и фиксированные эффекты будут идентичны, когда $T = 2$:

$$y_{i2} - \bar{y}_i = y_{i2} - \frac{1}{2}(y_{i2} + y_{i1}) = \frac{1}{2}(y_{i2} - y_{i1}) = \frac{1}{2}\Delta y_{i2},$$

и

$$y_{i1} - \bar{y}_i = y_{i1} - \frac{1}{2}(y_{i2} + y_{i1}) = -\frac{1}{2}(y_{i2} - y_{i1}).$$

Таким образом, после удаления временных средних мы записываем два уравнения, которые идентичны формулировке первых разностей.

Когда $T = 2$, FD довольно легко выполнить в любом программном пакете, который обеспечивает базовые преобразования данных, и легко вычислить статистику, основанную на гетероскедастичности, после FD-оценки (потому что в этом случае оценка FD — это просто пространственная регрессия).

Когда $T \geq 3$, FE- и FD-оценки не являются эквивалентными. Поскольку обе оценки являются объективными, мы не можем использовать объективность в качестве критерия. Кроме того, они обе состоятельны (при фиксированном T и $N \rightarrow \infty$).

При больших N и малых T выбор между FE и FD зависит от относительной эффективности оценок и определяется последовательной корреляцией, особенно в идиосинкразических ошибках u_{it} . (Мы будем предполагать гомоскедастичность u_{it} , так как сравнение эффективности требует гомоскедастических ошибок.)

Когда u_{it} последовательно некоррелированы, фиксированные эффекты более эффективны, чем первые разности. Поскольку обычно исследуется модель ненаблюдаемых эффектов (иногда неявно) с последовательно некоррелированными идиосинкразическими ошибками, оценка FE используется более часто, чем оценка FD. Но мы должны помнить, что это предположение может быть лож-

ным. Во многих приложениях можно ожидать, что ненаблюдаемые факторы, которые со временем меняются, будут последовательно коррелированы. Если u_{it} следует случайному блужданию, а это означает, что наблюдается очень существенная положительная последовательная корреляция, тогда разность Δu_{it} является последовательно некоррелированной и первые разности лучше. Во многих случаях u_{it} показывает некоторую положительную последовательную корреляцию, но, возможно, не такую высокую, как случайное блуждание. Поэтому мы не можем легко сравнить эффективность оценок FE и FD.

Трудно проверить, будут ли u_{it} серийно некоррелированы после оценки FE: мы можем оценить ошибки \hat{u}_{it} , но не u_{it} .

Часто бывает полезно попробовать как FE, так и FD: если результаты не чувствительны, тем лучше.

Когда T велико, и особенно когда N не очень большое (например, $N = 20$ и $T = 30$), мы должны быть очень осторожны при использовании оценки фиксированных эффектов.

Теоретически оценка фиксированных эффектов лучше (в смысле предоставления более точных оценок параметров, т. е. более маленьких стандартных ошибок), чем первых разностей, если ε_{it} последовательно некоррелированы.

С другой стороны, первые разности лучше, если $\Delta \varepsilon_{it}$ последовательно некоррелированы. Мы рассмотрели тесты и диагностическую статистику для проверки последовательной корреляции $\Delta \varepsilon_{it}$. Эти тесты могут помочь выбрать между FE и FD. Обратите внимание, что сложнее проверить последовательную корреляцию ε_{it} , потому что здесь нам нужно оценить все c_i .

Первые разности (оцененные по МНК) неприемлемы, если в модели имеется лаговая зависимая переменная. Другими словами, модель формы

$$y_{it} = \Phi y_{it-1} + \beta_1 x_{it1} + c_i + \varepsilon_{it}$$

не следует оценивать первыми разностями по методу МНК, поскольку Δy_{it-1} коррелирует с $\Delta \varepsilon_{it}$. В этом случае предпочтительной является оценка с фиксированными эффектами. Она будет смещена, когда

T мало, но это смещение исчезает со скоростью T^{-1} в выборках с большим количеством периодов времени.

Фиксированные эффекты легко реализуются не только для сбалансированных панелей, но и для несбалансированных панелей (несбалансированные панели представляют собой панели, в которых отдельные пространственные единицы могут иметь данные для другого количества периодов времени). Процедуры оценки фиксированных эффектов при несимметричной панели не намного сложнее, чем при сбалансированной панели. Если T_i — количество периодов времени для пространственной единицы i , мы просто используем эти наблюдения T_i , выполняя понижение размерности задачи. Тогда общее число наблюдений составит $T_1 + T_2 + \dots + T_N$. Как и в случае сбалансированной панели, одна степень свободы теряется для каждого наблюдения поперечного сечения из-за понижения размерности. Любой пакет регрессии с фиксированными эффектами делает соответствующую корректировку на данную потерю. Регрессия фиктивной переменной проходит точно так же, как при сбалансированной панели.

Оценка случайных эффектов (RE). Начнем с той же базовой модели с составной ошибкой:

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + c_i + \varepsilon_{it}.$$

Эта модель регрессии становится моделью случайных эффектов, когда мы предполагаем, что ненаблюдаемый эффект c_i не коррелирует с каждой объясняющей переменной:

$$\text{cov}(x_{ijt}, c_i) = 0, t = 1, 2, \dots, T, j = 1, 2, \dots, k.$$

Определим $v_{it} = c_i + \varepsilon_{it}$ как составную ошибку.

Ранее мы предполагали, что c_i коррелирует с x -ми, сейчас наоборот. В этом случае МНК будет состоятельным, но составная ошибка будет последовательно скоррелирована.

Предположим, что

$$E(\varepsilon_{it}) = 0, \text{ cov}(\varepsilon_{it}, \varepsilon_{is}) = 0, \text{ если } t \neq s,$$

и

$$\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0, \text{ если } i \neq j.$$

Для c_i $E(c_i) = 0$, $\text{cov}(c_i, c_j) = 0$, если $i \neq j$ и $\text{cov}(c_i, \varepsilon_{jt}) = 0$ для всех i, j, t и, самое главное, $\text{cov}(c_i, x_{it,j}) = 0$.

Тогда

$$E(v_{it} v_{is}) = \begin{cases} \sigma_c^2, & \text{для } t \neq s \\ \sigma_c^2 + \sigma_u^2, & \text{для } t = s \end{cases}.$$

Преобразуем модель и запустим GLS (обобщенный метод наименьших квадратов), чтобы решить проблему и сделать правильные выводы.

Основная идея состоит в том, чтобы сделать квазидифференцирование и удалить эффект последовательной корреляции. Расчеты для получения требуемого преобразования более продвинуты (матричная алгебра), но полученная оценка доступна в статистическом программном пакете Stata.

В конечном итоге со средневзвешенным значением МНК и фиксированных эффектов используйте квазипониженные данные:

$$\lambda = 1 - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_c^2},$$

$$y_{it} - \lambda \bar{y}_i = \beta_0 (1 - \lambda) + \beta_1 (x_{it,1} - \lambda \bar{x}_{i,1}) + \dots + \beta_k (x_{it,k} - \lambda \bar{x}_{i,k}) + (v_{it} - \lambda \bar{v}_i).$$

Если $\lambda = 1$, то это оценка фиксированных эффектов.

Если $\lambda = 0$, то это оценка МНК.

Итак, чем больше дисперсия ненаблюдаемого эффекта, тем ближе оценка к РЕ. Чем меньше дисперсия ненаблюдаемого эффекта, тем ближе к МНК.

Случайные эффекты vs фиксированные эффекты. Принято думать, что нужны фиксированные эффекты, поскольку считается, что проблема в том, что что-то ненаблюдаемое коррелирует с x -ми.

Одним из преимуществ оценки случайных эффектов является то, что она позволяет включать регрессионные системы, которые фиксируются во времени. Сюда относятся такие переменные, как образование. Однако, поскольку считается, что образование связано с ненаблюдаемой способностью, предположения о случайных эффектах могут быть в конце концов не такими привлекательными.

Если предположение о том, что c_i не коррелировано с x_{it} , выполняется, тогда FE-оценка валидная, но менее эффективная, чем оценка RE. Другими словами, если мы уверены, что случайные эффекты уместны, лучше использовать оценку RE.

Если нам действительно нужны случайные эффекты, единственная проблема — стандартные ошибки. Это означает, что можно запустить объединенную регрессию (а не преобразовывать случайные эффекты) и исправить только стандартную ошибку. Результирующая оценка менее эффективна, чем оценка случайных эффектов, но будет иметь правильные стандартные ошибки. Можно настроить стандартные ошибки для корреляции внутри группы. Также можно проверить, подходит ли спецификация случайных эффектов. Это делается с помощью теста Хаусмана. Идея состоит в том, что оценки с фиксированными эффектами валидны как при корреляции c_i с x_{it} , так и при ее отсутствии. Оценка со случайными эффектами, с другой стороны, валидна только в том случае, если c_i не коррелирует с x_{it} . В последнем случае она более эффективна, чем оценка с фиксированными эффектами.

Тест Хаусмана используется, когда нужно выбирать между оценкой со случайными эффектами и оценкой с фиксированными эффектами. Идея состоит в том, что мы должны использовать оценки случайных эффектов, если нулевая гипотеза по критерию Хаусмана не отклоняется:

$$\text{cov}(x_{ijt}, a_i) = 0, t = 1, 2, \dots, T, j = 1, 2, \dots, k.$$

На практике отказ отклонить означает, что либо оценки RE и FE достаточно близки, так что не имеет значения, какая используется, либо вариации выборки настолько велики в оценках FE, что невозможно сделать практически значимые различия статистически значимыми.

В последнем случае остается задаться вопросом, достаточно ли информации в данных для точной оценки коэффициентов. Отклонение с использованием теста Хаусмана означает, что ключевое условие RE неверно и используются оценки FE. (Естественно, как и во всех применениях статистического вывода, следует различать практически значимую разницу и статистически значимую разницу.)

Рассмотрим процедуру теста Хаусмана. Формулируются следующие гипотезы:

$$H_0: \text{cov}(\alpha_i, x_{it}) = 0,$$

$$H_a: \text{cov}(\alpha_i, x_{it}) \neq 0.$$

Если мы не можем отклонить H_0 , тогда оценки $\hat{\beta}_{FE}$ и $\hat{\beta}_{RE}$ не должны значительно отличаться. Тест проверяет, существует ли значительное отличие величины $(\hat{\beta}_{FE} - \hat{\beta}_{FE})$ от нуля (табл. 6).

Таблица 6

FE- и RE-оценки

Гипотеза	$\hat{\beta}_{FE}$	$\hat{\beta}_{RE}$
$H_0: \text{cov}(\alpha_i, x_{it}) = 0$	Состоятельная, неэффективная	Состоятельная, эффективная
$H_a: \text{cov}(\alpha_i, x_{it}) \neq 0$	Состоятельная	Несостоятельная

Если не можем отклонить H_0 , то используем RE.

Если H_0 отклоняется, то используем FE.

На практике требуется знать, какова $\text{cov}(\alpha_i, x_{it})$ для выполнения теста Хаусмана.

Статистика для проверки:

$$(\hat{\beta}_{FE} - \hat{\beta}_{RE})^T \left(\widehat{\text{var}}(\hat{\beta}_{FE}) - \widehat{\text{var}}(\hat{\beta}_{RE}) \right)^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}).$$

ГЛАВА 5. МОДЕЛИ БИНАРНОГО ВЫБОРА

Линейная модель вероятности

Бинарная случайная переменная представляет собой дискретную случайную переменную, которая имеет только два возможных значения: например, умирает ли объект (событие) или живет (несобытие). Такие события часто описываются как успех по сравнению с неудачей и кодируются с использованием значений 0 или 1. Следовательно, предположение о том, что этот тип переменной выбора имеет нормальное распределение, больше не выполняется. Наиболее распространенным распределением, используемым для бинарного выбора, является распределение Бернулли, которое принимает значение 1 с вероятностью успеха p и значение 0 с вероятностью провала $q = 1 - p$. Нахождение распределения для переменной выбора не является фиксированным. Например, если событие происходит очень редко, можно использовать распределение Пуассона.

Предположим, есть модель

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon,$$

где

$$y = \begin{cases} 1 \\ 0 \end{cases}$$

$$E(u | x_1, \dots, x_k) = 0.$$

Интерпретация β_j : изменение в $\Pr(y = 1 | x)$, дающее увеличение на одну единицу x_j при фиксировании других факторов.

$$E(u | x) = \Pr(y = 1 | x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

$$\Delta \Pr(y = 1 | x) = \beta_j \Delta x_j.$$

Пример (вероятность принятия на работу):

$$y = \beta_0 + \beta_1 netincome + \beta_2 educ + \beta_3 youngkids + \varepsilon,$$

где

$$y = \begin{cases} 1, & \text{если лицо работает} \\ 0, & \text{иначе} \end{cases}$$

Интерпретация коэффициентов модели:

β_1 : увеличение чистой прибыли на одну единицу изменяет вероятность быть нанятым на β_1 при фиксировании других факторов.

β_2 : увеличение времени обучения на один год изменяет вероятность быть нанятым на β_2 при фиксировании других факторов.

β_3 : увеличение числа маленьких детей на одного изменяет вероятность быть нанятым на β_3 при фиксировании других факторов.

Прогнозируемая вероятность < 0 или > 1 для определенных комбинаций значений для независимых переменных x .

Вероятность не может быть линейно связана с независимыми переменными для всех их возможных значений (например, появление первого ребенка уменьшает вероятность участия родителя в рабочем процессе, но появление последующих детей оказывает меньший предельный эффект).

Также необходимо исправить гетероскедастичность:

$$\text{var}(\varepsilon) = E(y_i)[1 - E(y_i)],$$

$$y_i = \beta^T x_i + \varepsilon_i, E(\varepsilon_i) = 0.$$

Поскольку y_i принимает значение 1 или 0, остатки ε_i могут принимать только два значения: $(1 - \beta^T x_i)$ и $(-\beta^T x_i)$. Соответствующие вероятности этих событий: $(\beta^T x_i)$ и $(1 - \beta^T x_i)$ (табл. 7).

Таблица 7

Значения $f(\varepsilon_i)$ для разных ε_i	
ε_i	$f(\varepsilon_i)$
$1 - \beta^T x_i$	$\beta^T x_i$
$-\beta^T x_i$	$1 - \beta^T x_i$

$$\begin{aligned}
\text{var}(\varepsilon_i) &= E(\varepsilon_i^2) - [E(\varepsilon)]^2 = \\
&= E(\varepsilon_i^2) = (1 - \beta^T x_i)^2 \beta^T x_i + (-\beta^T x_i)^2 (1 - \beta^T x_i) = \\
&= \beta^T x_i (1 - \beta^T x_i) = E(y_i) [1 - E(y_i)].
\end{aligned}$$

Оценим модель по МНК, получим \hat{y}_i .

Вычислим:

$$w_i = \sqrt{\hat{y}_i (1 - \hat{y}_i)},$$

$$\tilde{y}_i = \frac{y_i}{w_i},$$

$$\tilde{x}_i = \frac{x_i}{w_i} \quad (\text{для всех } x).$$

И, наконец, нам нужно запустить регрессию \tilde{y}_i на $\tilde{x}_{i1}, \tilde{x}_{i2}, \dots, \tilde{x}_{ik}$.

После всех этих шагов гетероскедастичность в модели корректируется.

Logit- и probit-модели

Существует модель

$$y_i^* = \beta^T x_i + \varepsilon_i,$$

y_i^* не наблюдаем:

$$y = \begin{cases} 1, & \text{if } y_i^* > 0 \\ 0, & \text{if } y_i^* \leq 0 \end{cases}$$

$$E(y_i^* | x_i) = \beta^T x_i,$$

$$\Pr(y_i = 1 | x) = \Pr(\varepsilon_i > -\beta^T x_i) = 1 - F(-\beta^T x_i),$$

$$0 < F(-\beta^T x_i) < 1.$$

F является совокупным распределением для ε_i .

В зависимости от вида функции распределения мы получаем логит- или пробит-модель.

Если функция кумулятивного распределения логистична, то мы имеем логит-модель:

$$F(-\beta^T x_i) = \frac{\exp(-\beta^T x_i)}{1 + \exp(-\beta^T x_i)} = \frac{1}{1 + \exp(\beta^T x_i)}.$$

Если функция кумулятивного распределения нормальна, то мы имеем пробит-модель:

$$F(-\beta^T x_i) = \int_{-\infty}^{\frac{-\beta^T x_i}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

Поскольку логистические и нормальные функции кумулятивного распределения близки друг к другу, за исключением хвостов, логит- и пробит-модели, скорее всего, дадут аналогичные результаты, если только выборки небольшие (так что у нас достаточно наблюдений у хвостов).

Однако оценки β из моделей логит и пробит напрямую не сопоставимы. Поскольку распределение логистики имеет дисперсию $\pi^2/3$,

$\hat{\beta}_{probit} = \frac{\sqrt{3}}{\pi} \hat{\beta}_{logit}$ (если $\sigma = 1$, т. е. стандартное нормальное распределение).

Альтернативный способ сравнения моделей:

1) рассчитайте сумму квадратов отклонений от прогнозируемых вероятностей; сравните правильно предсказанные проценты.

2) посмотрите на производные вероятностей относительно некоторой независимой переменной.

Обратимся к модели, описанной ранее:

— $x_{ik} \dots k^{th}$ — элемент вектора объясняющих переменных x_i ;

— $\beta_k \dots k^{th}$ — элемент β .

Производные вероятностей:

— линейная модель вероятности (LPM):

$$\frac{\partial}{\partial x_{ik}} (\beta^T x_i) = \beta_k;$$

— probit:

$$\frac{\partial}{\partial x_{ik}} \Phi(\beta^T x_i) = \phi(\beta^T x_i) \beta_k;$$

— logit:

$$\frac{\partial}{\partial x_{ik}} \Lambda(\beta^T x_i) = \frac{\exp(-\beta^T x_i)}{[1 + \exp(-\beta^T x_i)]^2} \beta_k.$$

В случае LPM производные являются постоянными.

В случае пробит- и логит-моделей нам нужно рассчитать их на разных уровнях объясняющих переменных, чтобы иметь представление о диапазоне вариаций полученных изменений вероятностей. Теоретически то, какой вариант лучше, зависит от процесса генерации неизвестных данных.

Ключевой выбор: $p_i = F(x_i, \beta)$.

В отличие от других приложений ML распределение определяется исключительно p_i , так что это единственная возможная ошибка спецификации.

Если $F(x_i, \beta)$ с ошибкой спецификации, то ММП (метод максимального правдоподобия, MLE) не состоятелен.

Но при условии, что p_i все еще имеет одноиндексную форму $F(x_i, \beta)$, выбор неправильной функции F влияет на все параметры наклона в равной степени, а отношение параметров наклона постоянное по моделям.

Модель logit — это бинарная модель, наиболее часто используемая статистиками. Logit просто обобщает многомиллионные данные (более двух выборов).

Модель probit — это бинарная модель, наиболее часто используемая экономистами. Probit обусловлена латентной нормальной случайной величиной. Probit распространяется на тобит-модели и мультиномиальный пробит.

Эмпирически можно использовать как логит-, так и пробит-модели:

- а) они обе дают аналогичные прогнозы и предельные эффекты;
- б) наибольшая разница заключается в предсказании вероятностей, близких к 0 или 1.

Дополнительная log-odds модель также является опцией, когда большинство выборов равно 0 или 1. МНК может быть полезен для предварительного анализа данных, но для получения конечных результатов должны использоваться методы пробит или логит.

ММП-оценка для logit. Функция правдоподобия:

$$L = \prod_{y_i=1} F(-\beta^T x_i) \prod_{y_i=1} [1 - F(-\beta^T x_i)] = \\ = \prod_{i=1}^n [F(-\beta^T x_i)]^{1-y_i} [1 - F(-\beta^T x_i)]^{y_i}.$$

Для оценки логит надо заменить:

$$F(-\beta^T x_i) = \frac{\exp(-\beta^T x_i)}{1 + \exp(-\beta^T x_i)} = \frac{1}{1 + \exp(\beta^T x_i)}.$$

Для пробит-оценки нужно заменить:

$$F(-\beta^T x_i) = \int_{-\infty}^{\frac{-\beta^T x_i}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt.$$

ММП-оценка для logit:

$$L = \prod_{i=1}^n [F(-\beta^T x_i)]^{1-y_i} [1 - F(-\beta^T x_i)]^{y_i} = \\ = \prod_{i=1}^n \left[\frac{1}{1 + \exp(\beta^T x_i)} \right]^{1-y_i} \left[\frac{\exp(\beta^T x_i)}{1 + \exp(\beta^T x_i)} \right]^{y_i} = \frac{\exp(\beta^T) \sum_{i=1}^n x_i y_i}{\prod_{i=1}^n [1 + \exp(\beta^T x_i)]}.$$

Обозначим:

$$t^* = \sum_{i=1}^n x_i y_i,$$

$$\log L = \beta^T t^* - \sum_{i=1}^n \log [1 + \exp(\beta^T x_i)],$$

$$s(\beta) = \frac{\partial(\log L)}{\partial \beta} = - \sum_{i=1}^n \frac{\exp(\beta^T x_i)}{1 + \exp(\beta^T x_i)} x_i + t^*.$$

Решается для β с использованием итерационного метода Ньютона — Рафсона.

П р и м е ч а н и е. Метод итерации Ньютона — Рафсона был применен для поиска последовательно лучших приближений к корням (или нулям) действительной функции. Идея метода заключается в следующем: действие начинается с первоначального приближительного подсчета, который достаточно близок к истинному корню, тогда функция аппроксимируется по касательной линии (которая может быть вычислена с использованием инструментов исчисления), и вычисляется x — точка пересечения этой касательной линии (благодаря элементарной алгебре). Данная точка пересечения x обычно будет лучшим приближением к корню функции, чем исходное предположение, и метод может быть повторен.

ММП-оценка для probit

$$\begin{aligned} L &= \prod_{i=1}^n [F(-\beta^T x_i)]^{1-y_i} [1 - F(-\beta^T x_i)]^{y_i} = \\ &= \prod_{i=1}^n [\Phi(-\beta^T x_i)]^{1-y_i} [1 - \Phi(\beta^T x_i)]^{y_i} = \\ &= \prod_{i=1}^n [\Phi(\beta^T x_i)]^{y_i} [1 - \Phi(\beta^T x_i)]^{1-y_i}, \end{aligned}$$

$$\log L = \sum_{i=1}^n y_i \log \Phi(\beta^T x_i) + \sum_{i=1}^n (1 - y_i) \log [1 - \Phi(\beta^T x_i)],$$

$$s(\beta) = \frac{\partial(\log L)}{\partial \beta} = - \sum_{i=1}^n \frac{y_i - \Phi(\beta^T x_i)}{\Phi(\beta^T x_i) [1 - \Phi(\beta^T x_i)]} \Phi(\beta^T x_i) x_i = 0.$$

Также решается для β с использованием итерационного метода Ньютона — Рафсона.

ГЛАВА 6. МУЛЬТИНОМИАЛЬНЫЕ И УПОРЯДОЧЕННЫЕ МОДЕЛИ

Мультиномиальные модели

Мультиномиальные и упорядоченные регрессионные модели — это простое расширение биномиальных моделей логистической регрессии. Они используются, когда зависимая переменная имеет больше двух значений.

Если зависимая переменная неупорядоченная, используется мультиномиальная модель. В противном случае, если зависимая переменная упорядочена, используется *упорядоченная модель*.

Пример неупорядоченных зависимых переменных:
уровень образования:

- $y = 1$, если у человека нет образования;
- $y = 2$, если человек имеет только общее образование;
- $y = 3$, если человек окончил колледж;
- $y = 4$, если человек окончил университет;
- $y = 5$, если у человека есть несколько высших образований.

Если у нас есть неупорядоченные переменные, нам нужно использовать мультиномиальные модели.

Но сложно тестировать мультиномиальную пробит-модель в предположении, что $\varepsilon_i \sim N(0,1)$. Поэтому существует мультиномиальная логит-модель, как правило, используемая в эконометрике в предположении, что $\varepsilon_i \sim e^{-e^{-t}}$.

Анализируется одновременный выбор среди m альтернатив. Идея состоит в том, чтобы взглянуть на парные сравнения с некоторым исходным результатом.

Пусть p_1, p_2, \dots, p_m — вероятности, связанные с указанными m -категориями.

$$\frac{p_j}{p_j + p_m} = F(\beta_j^T x), j = 1, \dots, (m-1),$$

$$\frac{p_j}{p_m} = \frac{F(\beta_j^T x)}{1 - F(\beta_j^T x)} = G(\beta_j^T x),$$

$$p_j = \frac{G(\beta_j^T x)}{1 + \sum_{j=1}^{m-1} G(\beta_j^T x)}.$$

Примечание. В примере бинарного выбора ($m = 2$) мы также начинали с определения:

$$\frac{p_j}{p_j + p_m} = \frac{p}{p + 1 - p} = p = F(\beta^T x),$$

$$p_{ij} = \frac{e^{x_i^T \beta_j}}{1 + \sum_{i=1}^{m-1} e^{x_i^T \beta_j}} \text{ для } j = 1, 2, \dots, (m-1),$$

$$p_{im} = \frac{1}{1 + \sum_{i=1}^{m-1} e^{x_i^T \beta_j}},$$

$$\sum_{l=1}^m p_{il} = 1.$$

Модель:

$$\ln \left(\frac{p_{ij}}{p_{im}} \right) = x_i^T \beta_j.$$

Мультиномиальный logit: MLE.

$$L = \prod_{i=1}^n p_{i1}^{y_{i1}} p_{i2}^{y_{i2}} \dots p_{im}^{y_{im}},$$

$$\log L = \sum_{i=1}^n \sum_{j=1}^m y_{ij} \log p_{ij},$$

$$y_{ij} = \begin{cases} 1, & \text{если } i\text{-й индивид выбирает } j\text{-ю категорию} \\ 0, & \text{иначе} \end{cases}$$

$$\sum_{j=1}^m y_{ij} = 1,$$

$$\frac{\partial(\log L)}{\partial \beta_k} = \sum_{i=1}^n (y_{ik} - p_{ik}) x_i = 0.$$

Решается для β с использованием итерационного метода Ньютона — Рафсона.

Упорядоченные модели

Предположим, что имеются 3 упорядоченные категории: 1, 2 и 3.

Индивид попадает:

- в категорию 3, если $\varepsilon < \beta^T x$;
- в категорию 2, если $\beta^T x < \varepsilon < \beta^T x + \mu$;
- в категорию 1, если $\varepsilon > \beta^T x + \mu$.

$$p_3 = F(\beta^T x),$$

$$p_2 = F(\beta^T x + \mu) - F(\beta^T x),$$

$$p_1 = 1 - F(\beta^T x + \mu).$$

Для функции распределения F можно использовать:

— логистическую функцию кумулятивного распределения, тогда это будет упорядоченная логит-модель:

— нормальную функцию кумулятивного распределения, тогда это будет упорядоченная пробит-модель.

Модель может быть расширена до m -категорий (добавление дополнительных пороговых значений μ).

Если имеется m -категория, то

$$p_m = F(\beta^T x),$$

$$p_{m-1} = F(\beta^T x + \mu_1) - F(\beta^T x),$$

$$p_{m-2} = F(\beta^T x + \mu_1 + \mu_2) - F(\beta^T x + \mu_1).$$

...

Это будет означать:

$$p_m = F(\beta^T x),$$

$$p_m + p_{m-1} = F(\beta^T x + \mu_1),$$

$$p_m + p_{m-1} + p_{m-2} = F(\beta^T x + \mu_1 + \mu_2),$$

$$p_m + p_{m-1} + p_{m-2} + \dots + p_2 = F(\beta^T x + \mu_1 + \mu_2 + \dots + \mu_{m-2}),$$

$$p_1 = 1 - F(\beta^T x + \mu_1 + \mu_2 + \dots + \mu_{m-2}).$$

Рассмотрим упорядоченный пробит с m -категориями. Определим набор констант: $\mu_1 = -\infty$, $\mu_m = \infty$, $\mu_1 < \mu_2 < \dots < \mu_m$.

$$Y_i^* = \beta^T x + \varepsilon_i \quad (i = 1, 2, \dots, n)$$

$$Y_i^* = \begin{cases} 1, & \text{если } Y^* \leq \mu_1 \\ 2, & \text{если } \mu_1 < Y^* \leq \mu_2 \\ \dots & \\ m, & \text{если } \mu_m < Y^* \end{cases}$$

Y^* не наблюдается, но мы знаем, к какой из m -категорий он принадлежит. Он относится к j -й категории, если

$$\mu_{j-1} < Y^* \leq \mu_j \quad (j = 1, 2, \dots, m).$$

Мы определяем набор фиктивных переменных:

$$z_{ij} = \begin{cases} 1, & \text{если } Y_i \text{ попадает в } j\text{-ю категорию} \\ 0, & \text{иначе} \end{cases}$$

$$\Pr(z_{ij} = 1) = \Phi(\mu_j - \beta^T x_i) - \Phi(\mu_{j-1} - \beta^T x_i).$$

MLE для упорядоченного probit:

$$L = \prod_{i=1}^n \prod_{j=1}^m [\Phi(\mu_j - \beta^T x_i) - \Phi(\mu_{j-1} - \beta^T x_i)]^{z_{ij}},$$

$$\log L = \sum_{i=1}^n \sum_{j=1}^m z_{ij} \log [\Phi(\mu_j - \beta^T x_i) - \Phi(\mu_{j-1} - \beta^T x_i)].$$

Ф.О.С. (условие первого порядка):

$$\frac{\partial(\log L)}{\partial \beta} = 0,$$

$$\frac{\partial(\log L)}{\partial \mu} = 0.$$

Решается для μ и β с использованием итерационного метода Ньютона — Рафсона.

Оценочная интерпретация в MNL и для упорядоченного логита, и для упорядоченного пробита:

— Для интерпретации как мультиномиальных, так и упорядоченных моделей предельные эффекты должны вычисляться для каждой категории y .

— В MNL мы рассматриваем значение коэффициентов и знаков.

— Интерпретация коэффициентов на основе их знака не очевидна в упорядоченной модели. Знаки параметров показывают, увеличивается ли латентная переменная y^* вместе с регрессором. Величина коэффициентов будет отличаться масштабным коэффициентом между пробит- и логит-моделями.

Предположим, что существует три категории.

Вероятности:

$$\Pr(y = 1 | x) = \Phi(-\beta^T x) = 1 - \Phi(\beta^T x),$$

$$\Pr(y = 2 | x) = \Phi(\mu - \beta^T x) - \Phi(-\beta^T x),$$

$$\Pr(y = 3 | x) = 1 - \Phi(\mu - \beta^T x).$$

Предельные эффекты:

$$\frac{\partial \Pr(y = 1 | x)}{\partial x} = -\phi(\beta^T x)\beta,$$

$$\frac{\partial \Pr(y = 2 | x)}{\partial x} = [-\phi(\beta^T x) - \phi(\mu - \beta^T x)]\beta,$$

$$\frac{\partial \Pr(y = 3 | x)}{\partial x} = \phi(\mu - \beta^T x)\beta.$$

Предполагая, что β положительна (для данного x), $\Pr(y = 1 | x)$ (нижняя категория) должна снижаться. Производная $\Pr(y = 1 | x)$ имеет противоположный знак от β .

Аналогично изменение в $\Pr(y = 3 | x)$ (высшая категория) должно иметь тот же знак, что и β . Но то, что происходит со средней ячейкой, неоднозначно. Это зависит от двух плотностей.

Предельный эффект от увеличения регрессора x зависит от вероятности выбора альтернативы j . Предельные эффекты каждой переменной зависят от разных альтернатив, суммированных до нуля. Каждое единичное увеличение в независимой переменной увеличивает (или уменьшает) вероятность выбора альтернативы j благодаря предельному эффекту, выраженному в процентах.

Цензурированные и усеченные модели регрессии

Иногда мы сохраняем нестандартные данные, которые содержат некоторые ограничения. Например, мы анализируем среднюю температуру пациентов в больнице: фактическая температура у них выше $36,6^\circ\text{C}$, но отчетная температура составляет $36,6^\circ\text{C}$. Другими словами, когда мы имеем дело с несовершенной информацией, мы можем использовать *цензурированные модели*. Цензурирование может быть вызвано проектированием опроса, случайной выборкой, когда мы знаем, где находятся пропущенные значения — выше или ниже заданного порогового значения, или если информация об объясняющих переменных x_i для всех i доступна, но для зависимой переменной y_i она отсутствует для некоторого i . Формально цензурированную модель можно сформулировать так:

$$y_i^* = \beta x_i' + \varepsilon_i, i = 1, 2, \dots, N,$$

$$y_i = y_i^*, \text{ если } y_i^* > 0,$$

$$y_i = 0, \text{ если } y_i^* \leq 0.$$

Модель также упоминается как модель цензурированной регрессии. Это стандартная регрессионная модель, где все отрицательные значения сведены к нулю. То есть наблюдения подвергаются цензурированию (снизу) в нуле. Когда данные подвергаются цензуриро-

ванию, изменение наблюдаемой переменной будет занижать эффект регрессоров на «истинную» зависимую переменную.

Таким образом, модель описывает две вещи. Одна из них — вероятность того, что $y_i = 0$ (заданный x_i):

$$\begin{aligned} P\{y_i = 0\} &= P\{y_i^* \leq 0\} = \\ &= P\{\varepsilon_i \leq -\beta x_i'\} = P\left\{\frac{\varepsilon_i}{\sigma} \leq -\frac{\beta x_i'}{\sigma}\right\} = \Phi\left(-\frac{\beta x_i'}{\sigma}\right) = 1 - \Phi\left(\frac{\beta x_i'}{\sigma}\right). \end{aligned}$$

Другая — распределение y_i , учитывая, что он положительный. Это усеченное нормальное распределение с ожиданием:

$$E\{y_i | y_i > 0\} = \beta x_i' + E\{\varepsilon_i | \varepsilon_i > -\beta x_i'\} = \beta x_i' + \sigma \frac{\phi(\beta x_i'/\sigma)}{\Phi(\beta x_i'/\sigma)}.$$

Итак, мы видим, что в тобит-модели моделируется экономическое поведение, которое часто дает нулевые результаты. В случае с цензурированной регрессией возникает проблема сбора данных.

$$y_i^* = \beta x_i' + \varepsilon_i, i = 1, 2, \dots, N,$$

$$y_i = y_i^*, \text{ если } y_i^* > c,$$

$$y_i = 0, \text{ если } y_i^* \leq c,$$

$$P\{y_i = 0\} = P\{y_i^* \leq c\} = 1 - \Phi\left(\frac{c - \beta x_i'}{\sigma}\right).$$

Модель цензурированной нормальной регрессии подвергается цензурированию справа или слева:

$$y_i^* = \beta x_i' + \varepsilon_i; \varepsilon_i | x, c_i \sim N(0, \sigma^2).$$

$w_i = \min(y_i, c_i)$: y_i наблюдается только в том случае, если оно меньше цензурирующего значения,

$w_i = \max(y_i, c_i)$: y_i наблюдается только в том случае, если оно больше цензурирующего значения, c_i .

Плотность w_i , заданная x и c_i :

$$f(w | c_i, x_i) = \begin{cases} 1 - \Phi\left(\frac{c_i - \beta x'_i}{\sigma}\right), w = c_i \\ \frac{1}{\sigma} \phi\left(\frac{w - \beta x'_i}{\sigma}\right), w < c_i \end{cases}$$

Логарифмическую вероятность для наблюдения i мы получаем путем взятия натурального логарифма плотности для каждого значения. Мы максимизируем сумму по i по отношению к β_j и s для получения MLE.

Как можно заключить из этой модели, зависимая переменная может подвергаться цензуре по разным причинам, однако можно добавить цензурированные наблюдения в регрессию. Также существует модель, где доступны только усеченные данные для регрессии.

Усеченная модель — это модель, в которой переменная наблюдается только тогда, когда зависимая переменная выше или ниже определенного порога. Здесь нет случайной выборки, но мы знаем правило, которое использовалось для включения единиц в выборку.

Усеченные модели используются для усечения с помощью исследования расчета или случайного усечения. Пример первого варианта: анализ силы в спорте, когда все спортсмены с силой ниже, чем 50 фунтов, отбрасываются из выборки. Пример второго варианта: только те, кто имеют работу, имеют информацию о заработной плате. Агенты определяют то, что не будет участвовать в исследовательском расчете, т.е. делают выборку.

То есть вместо использования всех n наблюдений мы используем подвыборку. Затем мы должны запустить МНК, используя только данную подвыборку (усеченную выборку). И что будет с МНК? Есть два случая:

1. МНК не смещен:
 - 1.1. Отбор выборки выполняется случайным образом;
 - 1.2. Отбор выборки определяется исключительно значением переменной x .
2. МНК смещен:
 - 2.1. Отбор выборки определяется значением переменной y ;

2.2. Отобранная выборка коррелирует с ε_i .

Второй случай невалидный, поэтому нам нужна усеченная регрессия с оценкой ML. Мы должны помнить, что $y_i < c_p$, поэтому находим функцию плотности ε_i при этом обстоятельстве:

$$\begin{aligned} f(\varepsilon_i | y_i < c_i) &= f(\varepsilon_i < c_i - x_i' \beta) = \\ &= \frac{f(\varepsilon_i)}{P(u_i < c_i - x_i' \beta)} = \frac{f(\varepsilon_i)}{P\left(\frac{\varepsilon_i}{\sigma} < \frac{c_i - x_i' \beta}{\sigma}\right)} = \frac{f(\varepsilon_i)}{\Phi\left(\frac{c_i - x_i' \beta}{\sigma}\right)} = \\ &= \frac{1}{\Phi\left(\frac{c_i - x_i' \beta}{\sigma}\right)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_i^2}{2\sigma^2}\right) = \frac{\frac{1}{\sigma} \varphi\left(\frac{\varepsilon_i}{\sigma}\right)}{\Phi\left(\frac{c_i - x_i' \beta}{\sigma}\right)}. \end{aligned}$$

Вклад функции правдоподобия для каждого i -го наблюдения определяется следующим образом:

$$L_i = \frac{\frac{1}{\sigma} \varphi\left(\frac{\varepsilon_i}{\sigma}\right)}{\Phi\left(\frac{x_i' \beta}{\sigma}\right)}.$$

Если мы берем логарифм, суммируем по i и максимизируем результат по отношению к β_j и σ^2 , то получаем MLE:

$$\begin{aligned} \log L(\beta, \sigma) &= \sum_{i=1}^N \log L_i = \\ &= -\frac{N}{2} \left[\log(2\pi) + \log(\sigma^2) \right] - \frac{1}{\sigma^2} \sum_{i=1}^N \varepsilon_i^2 - \sum_{i=1}^N \log \left[\Phi\left(\frac{x_i' \beta}{\sigma}\right) \right]. \end{aligned}$$

В итоге усеченная регрессия отличается от цензурированной регрессии следующим: зависимая переменная может быть подвергнута цензурированию, но вы можете включить цензурированные наблюдения в регрессию в модели с цензурированием и отбросить подмножество наблюдений. Таким образом, в усеченной модели только усеченные данные доступны для регрессии.

Раздел II

МЕТОДЫ МОНТЕ-КАРЛО

ГЛАВА 1. ВВЕДЕНИЕ В МЕТОДЫ МОНТЕ-КАРЛО

Методы Монте-Карло представляют собой группу аналитических подходов, основанных на генерации случайных чисел и использовании компьютерных алгоритмов. Эти методы широко применяются в вычислительных науках, а в последние десятилетия стали использоваться в социальных исследованиях.

Методы Монте-Карло существуют уже более 60 лет, они были одним из первых вычислительных подходов, который использовался на последних изобретенных компьютерах. Эти методы обычно применяются в тех случаях, когда традиционные подходы не дают точных результатов, будь то анализ сложных функций или моделирование рынков с большим количеством различных агентов. С развитием компьютерных технологий методы Монте-Карло нашли свое применение во многих дисциплинах, они остаются лидерами в эффективности и точности. Эти методы по-прежнему доминируют в своем первоначальном применении, которое имитирует сложные взаимодействия в любой области, где возможны количественные модели.

Наша цель — показать, что такое методы Монте-Карло и как их можно использовать для решения задач математики, экономики, физики, биологии, социологии и ряда других наук. Мы рассмотрим различные вычислительные подходы, использующие методы Мон-

те-Карло, и приведем примеры, иллюстрирующие, как эти методы могут быть использованы в практических исследованиях.

Методы Монте-Карло

По мере усложнения научных проблем и повышения мощности компьютеров появляются новые методы оценки и анализа. Методы Монте-Карло (ММК) являются их частью. Это методы анализа явлений, использующие вычислительные алгоритмы, основанные на генерации случайных чисел. Данная группа методов получила название от своих создателей — Станислава Улама и Джона фон Неймана, которые применили их во время работы над решением проблемы диффузии нейтронов в Лос-Аламосе в 1940-х гг. Название метода происходит от города Монте-Карло (Монако), известного своими казино, где дядя Улама любил играть в азартные игры, и отражает случайность, используемую и в методе, и в казино Монте-Карло.

Случайные числа часто используются в настоящее время в шифровании данных, например, в биткойн-транзакциях, платежах по кредитным картам, сетях TOR и I2P, кроме того, многие интернет-сервисы используют генерацию случайных чисел для защиты личных данных пользователей. Случайность также является ключевым элементом для всех видов игр, в которые включены события и взаимодействия. Методы Монте-Карло широко используются в науке, бизнесе, математике, логистике и многих других сферах. Эти методы позволяют создавать высоконадежные алгоритмы прогнозирования случайных процессов, например, движения частиц, логистики загруженного морского порта, передачи данных через Интернет и многие другие. Подход Монте-Карло также полезен для получения численных решений сложных задач, которые не могут быть решены аналитически.

Метод, который позже будет назван методом Монте-Карло, впервые использовал Жорж-Луи Леклерк, граф де Бюффон в 1777 г., предложив решение проблемы, которую он же выдвинул еще раньше, в 1733 г. Его эксперимент первоначально включал падение иглы на плоскую расчерченную поверхность и определение вероятности

пересечения иглой одной из линий. Это, в свою очередь, напрямую связано с оценкой π -значения с помощью случайных событий. Воспроизведем эту «проблему иглы» как введение в методы Монте-Карло.

Игла Бюффона

Рассмотрим иглу длины L , случайно брошенную на плоскую поверхность, ограниченную параллельными линиями на расстоянии d друг от друга. H_n является случайной величиной, представляющей количество раз, когда игла касается или пересекает линию в N бросков.

На рис. 20 x обозначает расстояние между центром иглы и ближайшей линией, а θ — острый угол между иглой и линией. В этой ситуации $0 \leq x \leq d/2$ и $0 \leq \theta \leq \pi/2$. Стрелка касается или пересекает линию тогда и только тогда, когда выполняется условие

$$x \leq \frac{L}{2} \sin \theta.$$

Если представить графически, то вероятность попадания иглы в линию равна площади под кривой $y = \frac{L}{2} \sin \theta$ (рис. 21). Итак, если $d \geq L$, то следующее верно:

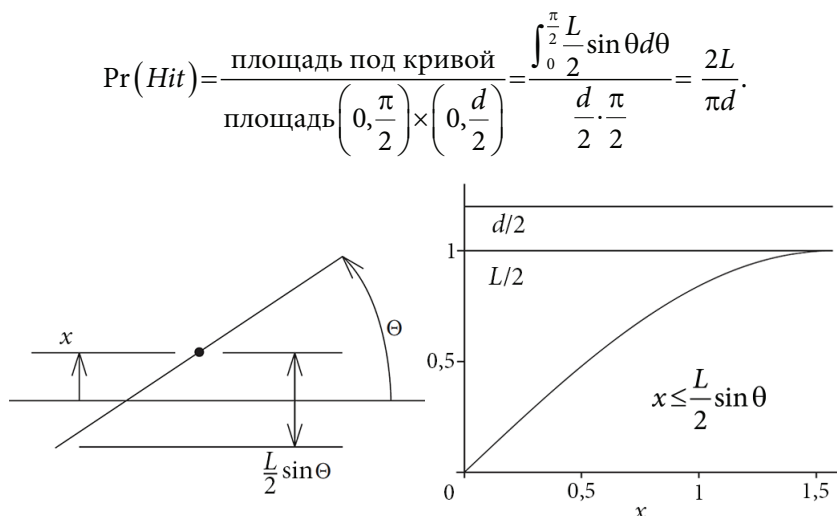


Рис. 20. Схема «проблемы иглы» Бюффона [3, с. 2]

Рис. 21. «Мишень» для иглы Бюффона [3, с. 3]

Если для простоты считать, что расстояние между линиями равно удвоенной длине иглы: $d = 2L$, то вероятность попадания будет равна $1/\pi$.

В реальном эксперименте величина π заранее неизвестна, и ее оценку можно получить из самого эксперимента. Поэтому Бюффон впервые нашел способ оценить известное π -значение с помощью случайных событий. Однако недостатком этого «инновационного» метода является то, что для достижения даже скромной точности требуется огромное количество таких случайных событий. На самом деле, чем больше экспериментов проводится, тем меньше точность: она определяется пропорцией $1/n$, где n — число экспериментов. Практика показала, что для получения достоверной оценки π необходимо провести более 500 экспериментов.

Первоначально моделирование производилось вручную с использованием механических калькуляторов. С развитием технологий моделирование стало проще. Джон фон Нейман, который был тесно связан с разработкой цифрового компьютера, понял, что эта технология может быть использована вместо ручного труда. Поэтому он разработал первый генератор случайных чисел — генератор «среднего квадрата», который, как упоминалось ранее, использовался для решения задачи потока нейтронов.

Чтобы проиллюстрировать, как программирование может быть использовано для генерации случайных чисел и решения математических задач, мы представим несколько программных кодов для MATLAB (или Octave) с результатами их выполнения. Каждый код имеет комментарии после знака %. Для получения дополнительной информации о командах, символах и функциях, используемых в кодах, см. файлы справки MATLAB или Octave. Во-первых, давайте представим и прокомментируем код для моделирования «проблемы иглы» Бюффона (код 1) и покажем гистограмму результатов (рис. 22).

Код 1. Симуляция проблемы иглы Бюффона + гистограмма полученных результатов (выполнено в MATLAB/Octave):

```
throws = 100;% 100 «бросков иглы»  
repl = 3000;% 3000 повторений
```



```

for i = 1: repl;% цикл повторений
x = rand(1, throws);% вектор 10000 псевдослучайных чисел в про-
межутке [0,1)
theta = 0.5*pi*rand(1, throws);% 10000 случайных чисел между 0
и  $\pi/2$ 
hits = x <= 0.5*sin(theta);% вектор попаданий, где 0 = промах 1 =
попадание
y(i) = sum(hits)/throws;% случайная переменная в числителе
end;
hist(y)% гистограмма
recipEst = sum(y)/repl% оценка для  $1/\pi$ 
vVec = (y-recipEst)*(y-recipEst);% вектор квадратичных отклоне-
ний
v = sum(vVec)/(repl-1);% вариация выборки
stddev = sqrt(v)% стандартное отклонение выборки
piEst = 1/recipEst% оценка  $\pi$ 

```

Скрипт выше (см. код 1) моделирует точечную оценку π и осно-
вывается на 3 тыс. повторений со 100 «бросками иглы», при усло-
вии, что расстояние между линиями равно удвоенной длине иглы

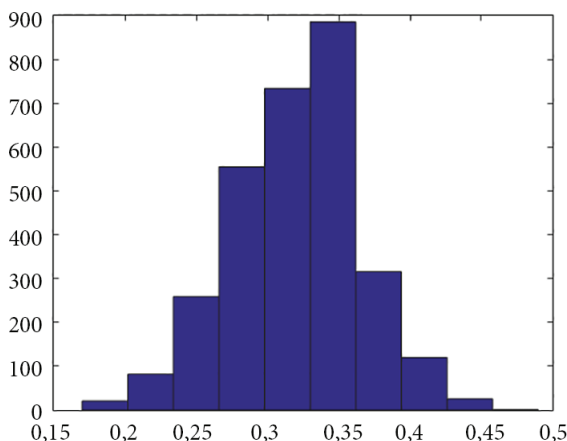


Рис. 22. Гистограмма иглы Бюффона для $1/\pi$

$d = 2L$. Он также вычисляет стандартное отклонение результатов и формирует гистограмму на $1/\pi$ (это, по сути, вероятность «попадания» при выбранных параметрах) (см. рис. 22). Программа генерирует два случайных числа для вычислительных целей, одно играет роль x , $0 \leq x < d/2$, а другое — θ , $0 \leq \theta < \pi/2$. Затем вычисляется, справедливо ли $x \leq \frac{L}{2} \sin \theta$, и отмечается попадание, если да. Используя повторения как «суперэксперименты», можно построить гистограмму по вычисленным данным.

Среднее из 3 тыс. суперэкспериментов аппроксимирует среднее значение распределения и является оценкой $1/\pi$. Для запуска скрипта, показанного на рис. 22, среднее значение $p = 0,3183$. Гистограмма считается очень полезным представлением результатов эксперимента Монте-Карло. Более того, анализ таких графиков позволяет лучше понять результаты эксперимента. Гистограмма представляет собой график, который показывает, сколько значений попадают в установленные интервалы. В зависимости от количества значений, лежащих в пределах определенного интервала, определяется высота столбика этого интервала. По сути, это называется частотной гистограммой. Деление частот на общее число значений позволяет оценить относительные частоты. Используя эти относительные частоты для гистограммы, мы построим гистограмму плотности, которая отображает вероятностное распределение значений экспериментов.

После запуска скрипта мы получили следующие оценки: $1/\pi = 0,3183$, $\pi = 3,1410$, и стандартное отклонение $s = 0,0460$. Одним из важных свойств нормального распределения является то, что выборка, имеющая такое распределение, с вероятностью 0,954 попадает между двумя стандартными отклонениями по обе стороны от среднего. Если количество выборок будет увеличено, то интервал будет сокращаться, соответственно, в нашем эксперименте интервал

при 3 тыс. выборок составил бы $\frac{2 \cdot 0,0460}{\sqrt{3000}} = 0,0016$ по обе стороны

от средней. Поэтому в нашем случае любая взятая выборка будет находиться внутри интервала (0,3151; 0,3215) с вероятностью 95 %. Можно также сказать, что π с вероятностью 95 % лежит в следующем

интервале: $1/0,3215 < \pi < 1/0,3151$. То есть в нашем примере π находится между 3,1104 и 3,1735.

Второй способ представления результатов эксперимента — построение траектории его перезапусков, представляющей собой два значения — номер эксперимента и его результат. В качестве иллюстрации рассмотрим эксперимент с азартным игроком и казино. Игрок делает ставки против казино с равными шансами на выигрыш. Игрок начинает со 100 долл., в то время как ставки казино значительно выше, например, 2000 долл., игра продолжается пока один из участников не обанкротится. Код 2 имитирует этот эксперимент. В данном эксперименте мы попробуем найти ответы на два вопроса:

- Может ли казино обанкротиться?
- Как долго продлится игра (или когда один из игроков обанкротится)?

Код 2. Эксперимент с игроком и казино (выполнено в MATLAB/Octave):

```
R = zeros(1,2000000); i = 1; R(i) = 100;% вектор 2,000,000 нулей
while(R(i) > 0 & R(i) < 2100)
    i = i+1;
    W = (rand < 0.5);% случайное значение 0 или 1
    W = 2*W — 1;% случайное значение +1 или -1
    R(i) = R(i-1)+W;% новое состояние игрока
end
subplot(1,2,1); plot(1: i+1, R(1: i+1))% построим график R и его
индекса
%%% нижеследующий код необходим для расчета времени игры
for j = 1:20 % цикл попыток
    i = 1; R(i) = 100;
    while(R(i) > 0 & R(i) < 2100)
        i = i+1;
        W = 2*(rand < 0.5)-1;
        R(i) = R(i-1)+W;
    end
    T(j) = i;% результат расчета времени игры
end
subplot(1,2,2); hist(T)% гистограмма результатов
```

Опять же в этом коде мы используем встроенный генератор случайных чисел, который возвращает значения в диапазоне $[0, 1)$. Предположим, что игрок выигрывает, если сгенерированное случайное число меньше 0,5, во всех остальных случаях он проигрывает.

На рис. 23, а представлена примерная динамика эксперимента для первой попытки, построенной средствами MATLAB. Как мы видим, игрок обанкротится менее чем за 8 тыс. ставок. Выполнение еще нескольких запусков первой части скрипта показывает, что ситуация, когда казино обанкротится, крайне маловероятна. На рис. 23, б представлен результат второй части кода, который возвращает гистограмму игрового времени на основе 20 последующих запусков. Следует отметить, что верхняя граница длительности игрового времени отсутствует, как видно из гистограммы, игра может занять очень много времени, в нашем эксперименте это 1400 тыс. ходов.

Поскольку мы обсуждаем основы методов Монте-Карло, чтобы полностью понять, как эти методы могут быть использованы на практике, нам необходимо рассмотреть две важные концепции — вероятности случайного события и генерации случайных чисел.

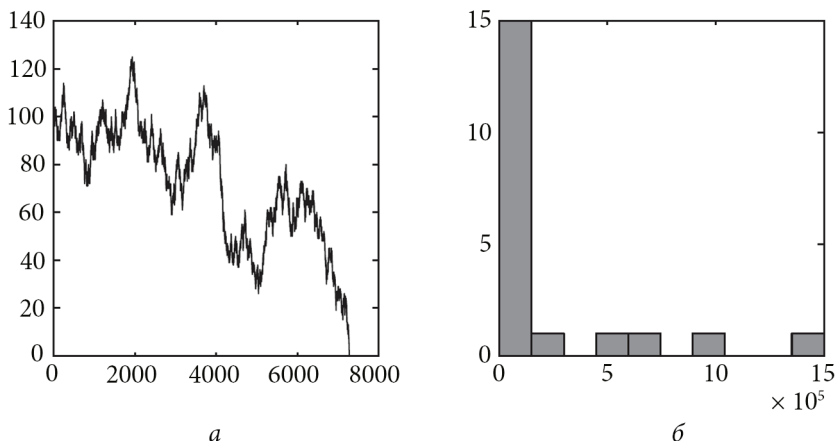


Рис. 23. Примерная динамика эксперимента:
а — денежное состояние игрока; б — гистограмма количества игр до банкротства в 20 экспериментах

Вероятность

Понимание концепции вероятности является очень важной частью понимания методов Монте-Карло. Мы покажем, что такое вероятность и как случайные величины могут быть использованы в оценках.

Рассмотрим эксперимент с событием E , имеющим множество всех возможных исходов Ω . Например, в эксперименте с подбрасыванием монеты два раза всего четыре возможных исхода. При каждом броске монета может приземлиться либо орлом (H), либо решкой (T), поэтому $\Omega = \{(H, H); (H, T); (T, H); (T, T)\}$. В этом эксперименте в качестве случайной величины X мы могли бы использовать ряд результатов «орел», поэтому X может быть 0, 1 или 2. Поскольку все исходы одинаково вероятны, их вероятности $\Pr(X)$, по существу, — это просто подсчет успешных исходов для соответствующего события. $\Pr(X = 2) = 1/4$, поскольку только результат (T, T) подходит, $\Pr(X = 1) = 1/2$, потому что (H, T) и (T, H) содержат одного «орла», $\Pr(X = 0) = 1/4$, поскольку только результат (H, H) подходит.

Давайте разделим результаты E на два набора: E_1 и E_2 , которые удовлетворяют двум важным условиям: $E = E_1 \cup E_2$ и $E_1 \cap E_2 = \emptyset$, другими словами, эти два множества называются непересекающимися. Поэтому для двух событий в целом

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2).$$

Вычитание важно, чтобы избежать двойного подсчета значений, которые появляются в обоих наборах. Для двух непересекающихся событий эта формула преобразуется в

$$\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2), \text{ если } E_1 \cap E_2 = \emptyset.$$

Дискретные и непрерывные случайные величины. Все случайные величины можно разделить на две основные группы: дискретные и непрерывные. *Дискретные* случайные величины — это те, которые имеют конечное число или бесконечно много результатов, поддающихся подсчету (очевидно, по одному).

Эксперимент с подбрасыванием монет, например, имел четыре результата, поэтому случайная величина для числа «орлов» попадает в группу дискретных случайных величин. Пример счетно-бесконечного числа исходов — эксперимент, где монета подбрасывается и выпадает «орел» впервые. В принципе для этого может потребоваться любое количество бросков монет, поэтому нельзя оценить случайную величину окончательно, но она по-прежнему дискретна.

Существует множество *непрерывных* процессов, к ним относятся такие явления, как течение времени, движение и многие другие. Очевидно, что измерения таких переменных ограничены конечными оценками, которые используются для описания непрерывных случайных величин. Формально непрерывные случайные величины являются переменными для $\Pr(X = x) = 0$, где x — действительное значение.

Еще одной важной вероятностной концепцией является кумулятивная функция распределения, описывающая вероятность случайной величины X с некоторым определенным распределением вероятностей, которое будет не больше x . Она представлена в следующей математической форме:

$$\text{cdf}(x) = \Pr(X \leq x).$$

Чтобы проиллюстрировать, что такое *кумулятивная функция распределения* (cdf), вернемся к эксперименту с двойным подбрасыванием монет и случайной переменной X , которую мы использовали выше. Существует несколько возможных значений x (рис. 24):

1. Если $x < 0$, то $\text{cdf}(x) = 0$, так как нет результатов, для которых $X < 0$.
2. При $x = 0$ cdf увеличивается до $1/4$, так как $\Pr(X = 0) = 1/4$.
3. Если $0 < x < 1$, то cdf не изменяется и поэтому остается в значении $1/4$ на этом интервале.
4. При $x = 1$ cdf снова поднимается, на этот раз на $1/2$. В этот момент $\text{cdf}(1) = 3/4$, так как событие $X \leq 1$ состоит из нескольких исходов: $\{(T, T); (H, T); (T, H)\}$.
5. В то время как $1 < x < 2$, cdf не изменяется и имеет фиксированное значение $3/4$, как в предыдущем шаге.

6. Наконец, при $x = 2$ cdf увеличивается на $1/4$ до 1 и остается там, так как теперь включены все возможные результаты.

В целом x , стремящийся к минус бесконечности, означает, что все больше и больше результатов эксперимента опущены, а cdf должен стремиться к 0.

$$\lim_{x \rightarrow -\infty} \text{cdf}(x) = 0.$$

С другой стороны, если x стремится к плюс бесконечности, то включается все больше и больше результатов, а cdf стремится к 1.

$$\lim_{x \rightarrow \infty} \text{cdf}(x) = 1.$$

Таким образом, также верно, что cdf увеличивается или, по крайней мере, остается таким же, если увеличивается x , потому что все больше и больше результатов удовлетворяют условию $X \leq x$. Иными словами, $\text{cdf}(x_1) \leq \text{cdf}(x_2)$ является верным, когда $x_1 \leq x_2$. Если случайная переменная дискретна, ее cdf будет меняться в определенных дискретных точках, соответствующих значениям x , для которых один или несколько результатов удовлетворяют $X = x$. В противном случае cdf остается постоянным. В качестве иллюстрации можно привести cdf для бросания пары кубиков (рис. 25). Очевидно, то же самое не будет справедливо для непрерывных случайных величин, где cdf — гладкая линия.

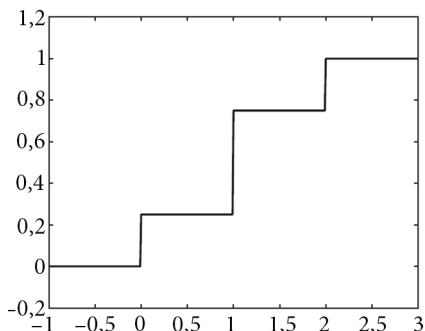


Рис. 24. Cdf по количеству «орлов»
(x против cdf) [3, с. 12]

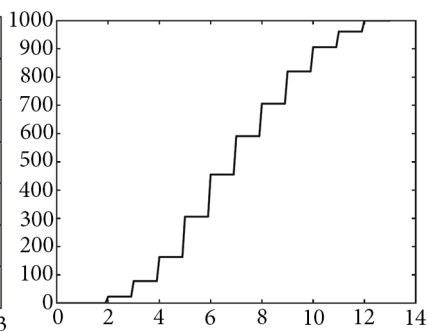


Рис. 25. Cdf для бросания кубиков
(x против cdf) [3, с. 13]

Как видим, случайные величины и вероятность имеют свои отличительные особенности, важные для их использования в вычислительных целях. Однако перед использованием случайные величины должны быть получены или сгенерированы каким-то образом. Далее рассмотрим это.

Генерация случайных чисел

Рассмотрим понятие генерации случайных чисел. Существует два основных метода генерации случайных чисел. Первый метод основан на измерениях определенного физического явления, которое считается случайным, например, теплового шума, атмосферного шума или любых квантовых явлений. Строго говоря, только этот метод может дать «истинно» случайные числа. Этот метод тем не менее трудно реализовать на практике, так как он довольно медленный и требует специального оборудования. Второй метод коррекции смещения основан на использовании определенных вычислительных алгоритмов. Поскольку компьютеры и алгоритмы генерации детерминированы, числа, полученные с помощью этого метода, не являются «истинно» случайными. Гипотетически, если исходные данные и алгоритм известны, то число можно вычислить с определенной легкостью. Тем не менее эти числа, как правило, оказываются сложнопредсказуемыми и обладают рядом признаков, позволяющих говорить о достаточном уровне «случайности». Числа, сгенерированные этим методом, называются «псевдослучайными». Говоря о случайных числах в этом пособии, мы, как правило, будем иметь в виду числа, рассчитанные именно вторым методом, так как его проще реализовать и проще использовать для моделирования методом Монте-Карло.

Существует несколько основных требований к генератору случайных чисел (ГСЧ). Приведем пять требований, наиболее часто встречающихся в литературе, посвященной методам Монте-Карло:

1. Быстрая скорость. Программы, использующие генераторы случайных чисел, обычно требуют миллионы случайных чисел

за короткое время. Таким образом, генератор должен уметь быстро вычислять случайные числа.

2. Повторяемость. Это требование необходимо для отладки. Если алгоритм не может быть повторен, то его невозможно улучшить или исправить.

3. Анализируемость. Для обеспечения распределительных свойств полученных значений результаты генерации должны быть по возможности проанализированы.

4. Длительный период. В конце концов, все алгоритмические ГСЧ повторяются после определенной точки генерации, поэтому чем дольше период до этой точки, тем лучше.

5. Очевидная случайность. Как уже упоминалось ранее, алгоритмы не позволяют генерировать действительно случайные числа, поэтому, по крайней мере, эти числа должны казаться случайными для предполагаемой сферы использования.

Поначалу кажется, что для генерации случайного числа достаточно использовать сложный алгоритм с множеством шагов и преобразований. Это не соответствует действительности.

Общая форма генератора случайных чисел будет следующей:

$$\text{Output} = f(\text{variables}, \text{parameters}).$$

Для некоторой функции f для каждого запроса программа выполняет алгоритм вычисления этой функции, используя переменные и параметры для получения определенной выборки «случайного» процесса. Параметры функций являются фиксированными значениями, а переменные изменяются от примера к примеру. Переменными могут быть любые внешние значения, такие как время, или внутренние значения, которые изменяются на каждой итерации. Эти внутренние переменные принято назвать *семенами* (*seed*). Генератор может иметь одно такое значение или несколько. Сначала рассмотрим генераторы одиночных семян.

Генераторы с одним сохраненным значением. Если базовые значения были взяты из какого-либо внешнего процесса в любой момент во время выполнения алгоритма генерации, то нельзя будет сгенерировать точно такие же примеры еще раз, в этом случае невозможно отладить программу. Также было бы трудно обеспечить

распределительные свойства. Поэтому обычно начальные значения генерируются самим ГСЧ, за исключением первого, которое, например, может быть взято из компьютерных часов и сохранено для отладки.

После начального введения базового числа все остальные базовые значения вычисляются из самой функции f и одновременно вычисляются выходные выборки. Процесс может быть упрощен, если в качестве следующего базового значения используется выходная выборка. Это не приведет к необъективным результатам, так как алгоритм все равно скорректирует выборки.

Как уже упоминалось ранее, последовательности исходов в любом генераторе случайных чисел имеют тенденцию повторяться через некоторое время. Это происходит после того, как повторяется базовое число. При едином базовом числе ГСЧ период уникальной выходной последовательности, как правило, не превышает количества различных возможных базовых чисел.

В некотором смысле генератор случайных чисел похож на книгу с длинным списком чисел, а начальная переменная выбирает стартовую точку для чтения чисел, после чего числа следуют заданному порядку. Если используется другое начальное число, то алгоритм начинает «читать» с другого места. Если используется одно и то же начальное число, то последовательность выходов одинакова. Когда алгоритм достигает конца списка, то «чтение» снова начинается со случайного места в списке.

Этот тип генераторов быстр и относительно прост в проверке на ошибки, но, с другой стороны, период случайных чисел этого типа генераторов довольно ограничен. К счастью, другие типы генераторов не имеют такого недостатка.

Техника средних квадратов и других средних значений. В основе этого типа генераторов случайных чисел лежит идея, что после первых нескольких цифр в вычислении большинства математических функций, таких как $\sin()$ или $\log()$, следующие цифры практически случайны. Фон Нейман во время работы в Лос-Аламосе впервые применил этот тип ГСЧ, изначально он использовал x^2 в качестве ключевой функции, а затем предположил, что средние цифры переменной являются «случайными», что дало название типу.

Выбор функции был обусловлен тем, что $\sin()$ или $\log()$ являются относительно трудоемкими вычислениями даже при использовании современных технологий. Код 3 демонстрирует это.

Код 3. Сравнение среднезначного RNG и встроенного RNG (исполняемого в MATLAB / Octave)

```
x = 0.5;
tic
for i = 1:1000000
    x = 100000*sin(x);
    w = floor(x);% берем первые 5 цифр
    x = x-w;% оставшиеся цифры
end
toc% 0.0369 вычисление синуса потребует больше времени
% сравнение со встроенным генератором Matlab
tic
x = rand(1,1000000);
toc% 0.0139
```

Очевидно, что точные значения времени выполнения скрипта могут варьироваться в зависимости от вычислительной мощности используемой машины, но встроенный генератор случайных чисел всегда в 3–30 раз быстрее метода средних квадратов, хотя оба возвращают аналогичные псевдослучайные результаты.

Основным недостатком этого метода является короткий период и быстрый цикл для большинства исходных переменных. Кроме того, как показано выше, генераторы средних цифр работают медленно для многих программ. Поэтому данная методика имела ограниченное практическое применение. Это стимулировало поиск более подходящих генераторов случайных чисел.

Линейные конгруэнтные генераторы случайных чисел. Линейные конгруэнтные генераторы случайных чисел являются одними из наиболее широко используемых в настоящее время, они зависят от простого повторения для вычисления следующего результата в последовательности из предыдущего. Например, многие программные системы, такие как MATLAB, R, Python, PHP, Ruby и др.,

используют линейный конгруэнтный генератор случайных чисел в качестве стандартного ГСЧ, который обычно представлен в их документации.

Название генератора происходит от модульной арифметики, где a можно назвать конгруэнтным b по модулю m , если $m \mid (a - b)$. Формально определенный линейный конгруэнтный генератор случайных чисел (LCRNG) является функцией f в зависимости от трех фиксированных параметров a , c , m :

$$f(x) = (ax + c) \bmod m,$$

где x — целое.

Давайте посмотрим, как этот генератор случайных чисел работает на практике. Например, возьмем $a = 5$, $c = 1$, $m = 8$ и используем $x = 0$ в качестве начального числа. Когда мы используем алгоритм, выход будет $(0 + 1) \bmod 8 = 1$, и 1 будет следующим начальным числом. Если мы вычисляем функцию повторно, мы получим следующую последовательность выборок перед началом цикла вывода:

$$R = (0, 1, 6, 7, 4, 5, 2, 3, 0).$$

Этот выход никак нельзя назвать случайным, так как он генерируется простой формулой, но с разными значениями m и a результаты могут быть гораздо более случайными. Как видно, длина выходного периода (количество различных возвращаемых значений) этого генератора равна значению m , поэтому в любом реальном генераторе берется как можно большее значение m , например, упомянутый ранее алгоритм обычно имеет период длиной $2^{19937} - 1$, что обеспечивает более 10^{6000} уникальных исходов, пока они не станут однообразными. В целях оптимизации m можно принять за 2^k , где k равно количеству операций, которые процессор может обработать за один раз (обычно 32 или 64 для современных устройств). Этот метод достаточно удобен, так как операция по модулю не займет дополнительного времени и произойдет автоматически в связи с тем, что нет места для битов более высокого порядка. В этом случае LCRNG потребуется только сложение и умножение, чтобы получить один результат, что является быстрым способом и экономит

вычислительную мощность. Для чисел с плавающей запятой также требуется деление.

Если $c = 0$, то операция сложения избегается, и последовательность может быть представлена следующим уравнением:

$$x_{n+1} = ax_n \bmod m.$$

В этом случае максимальный период не может быть равен m , так как $x_n = 0$ также превращает все последующие значения в 0. Период $m - 1$ все еще может быть достигнут в ситуации, когда m — простое число, однако операция по модулю больше не будет автоматической, в то время как m не является степенью 2.

В общем, если $C = 0$, m не простое число, а если x_n является целочисленным множителем делителя d числа m , то все последующие значения x будут также кратны d . У нас будет $x_n = pd$ для некоторого целого числа p , и так как d делит m , то $m = qd$ для некоторого целого числа q . То есть $x_{n+1} = ax_n \bmod m = apd \bmod m$ означает $x_{n+1} - apd = rd$ для некоторого целого числа R . Но $x_{n+1} = (ap + rq)d$. Поэтому, когда $c = 0$, x в идеале должен проходить через все целые числа относительно простых до m , если a выбран правильно. Заметим, что если $c = 0$ и 1 в последовательности, то она будет состоять из последовательных степеней a .

Как уже упоминалось ранее, многие встроенные псевдо-ГСЧ линейно конгруэнтны. Однако в настоящее время все больше пакетов программ используют более совершенные генераторы.

При использовании линейных конгруэнтных генераторов случайных чисел можно столкнуться с рядом проблем. Во-первых, период ограничен не более чем m . Во-вторых, если выборки действительно случайны, повторы обычно возможны, но для LCRNG это невозможно. Также анализ последовательностей позволяет делать определенные прогнозы относительно следующих выходных значений, что крайне нежелательно для хорошего генератора случайных чисел. Основной проблемой LCRNG является их простота. Во избежание этого потенциально могут использоваться многомерные генераторы конгруэнтного типа и генераторы с другой математической основой.

Несколько применений методов Монте-Карло

Как уже упоминалось ранее, методы Монте-Карло могут быть использованы во многих областях математики, науки, техники или даже развлечений. Рассмотрим несколько вариантов применения методов Монте-Карло для решения математических или статистических задач.

Метод «дартса» (Hit-or-Miss). Одним из самых ранних применений метода Монте-Карло является оценка интегралов, так как некоторые из них могут быть трудны для решения «традиционным» способом. Для решения может быть использован метод Hit-or-Miss. Случайным образом предлагая числа, мы можем определить искомую площадь, пользуясь той же логикой, что и при моделировании иглы Бюффона.

Рассмотрим общую ситуацию. Пусть $\theta = \int_a^b f(x) dx$, что является областью под функцией f между $x = a$ и $x = b$, при условии, что $f(x) \geq 0$ на интервале $[a, b]$. Пусть $M \geq \max_{a \leq x \leq b} f(x)$, а целевая область лежит внутри прямоугольника $[a, b] \times [0, M]$. Если мы начнем выбирать точки внутри этого прямоугольника случайным образом, то вероятность попадания точки в целевую область равна отношению этой

области к площади прямоугольника: $\frac{\theta}{M(b-a)}$. Если n — количество бросков, а h — количество попаданий, то верно следующее:

$$\frac{h}{n} \approx \frac{\theta}{M(b-a)}$$

или

$$\theta \approx \left(\frac{h}{n}\right) M(b-a).$$

Таким образом, поскольку размеры прямоугольника, n и h известны, искомая площадь может быть легко вычислена простыми операциями, даже в случае, если целевая функция крайне сложная.

Используя ту же логику, методы Монте-Карло можно весьма эффективно применять для решения интегралов в множественном интегрировании по сложным областям в многомерном пространстве. Целевая область «заключается» в определенную легко вычисляемую площадь или пространство, а затем внутри нее случайным образом выбираются точки и проверяется их попадание в выбранное простое пространство. Коэффициент попадания «броска» будет оценкой искомой площади. Чтобы проиллюстрировать это, давайте еще раз оценим значение π с помощью инструментов MATLAB.

Мы берем кривую линию, образующую первый квадрант единичной окружности, $y = \sqrt{1 - x^2}$, $0 \leq x \leq 1$. Поэтому целевая область $\pi/4$ содержится в квадрате $[0, 1] \times [0, 1]$, который имеет площадь, равную 1. После применения метода Hit-or-Miss оценка π является следующей:

$$\pi \approx 4 \frac{h}{n}.$$

Приведем код MATLAB, который позволяет выполнить это моделирование и оценить, насколько хороша оценка π (код 4).

Первая часть кода моделирует два значения — количество попаданий и оценку π . В проведенном запуске значения были 7815 и 3,1260 соответственно, что достаточно точно даже для такого относительно небольшого количества испытаний. Очевидно, что увеличение числа попыток окажет положительное влияние на оценку. Продемонстрируем это.

Вторая часть кода позволяет увидеть, как оценочное значение π улучшается по мере увеличения количества испытаний. Всего было проведено 20 циклов с 1 048 576 испытаниями. На первых 5 циклах оценка значения π очень неточна, но после седьмого цикла все значения становятся близки к 3,1, а на последних циклах сходятся к 3,14, что можно назвать довольно точной оценкой (рис. 26), при этом абсолютная погрешность уменьшается в каждом новом цикле со скоростью примерно $1/\sqrt{n}$.

Код 4. Оценка π (выполнено в MATLAB/Octave)

```
nTrials = 10000;
```

```
x = rand(1, nTrials); % генерация x-координат
```

```

y = rand(1, nTrials);% генерация y-координат
hits = sum(x.^2 + y.^2 < 1)% 7815 бросков
piEst = 4*hits/nTrials% 3.1260
% пауза
nPower = 20;% определение периода ГСЧ
for i = 1: nPower% 2^20=1048576
x = rand(1,2^i);
y = rand(1,2^i);
piEst(i) = 4*sum(x.^2 + y.^2 < 1)/2^i
end
plot(piEst)% вывод графика оцененной сходимости  $\pi$ 

```

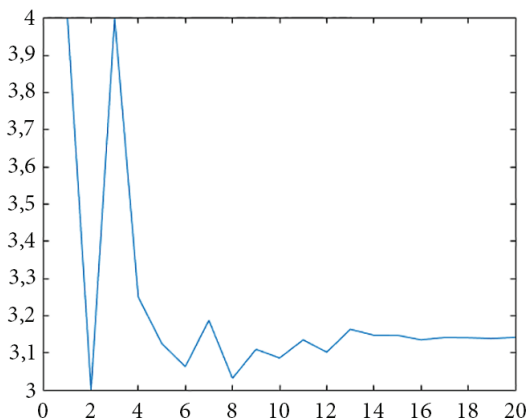


Рис. 26. Схождение оценки π
(оценка против количества попыток)

Как видим, методы Монте-Карло достаточно эффективны в данной ситуации и не требуют решения задачи вычисления интегралов.

ГЛАВА 2. СЭМПЛИРОВАНИЕ ДАННЫХ

В статистике, как правило, сэмплирование данных — это анализ, используемый для выбора данных и управления их репрезентативным набором с целью поиска закономерностей и тенденций. Сэмплирование позволяет исследователям работать с управляемым объемом данных и быстрее создавать и запускать модели. В настоящее время, как мы уже упоминали ранее, исследователи все больше используют вероятностные модели, поскольку для «традиционных» аналитических подходов данные задачи слишком сложные. Кроме того, вероятностные модели позволяют исследователям отказаться от нереалистичных предположений о нормальности и независимости, необходимых для аналитических методов.

Сэмплирование методом Монте-Карло используется для предсказания поведения моделей при заданном наборе условий или, например, для нахождения соответствующих значений параметров модели с заданным набором экспериментальных данных. Общий алгоритм здесь состоит в разделении задачи сэмплирования сложных распределений на несколько подзадач, которые решить уже проще. Мы рассмотрим два типа выборки для различных типов распределений: обратное преобразование и выборку отбраковки. Оба они хороши для ситуаций с однозначными результатами; однако существуют более сложные подходы (например, метод Монте-Карло по схеме марковской цепи), которые могут быть использованы с многомерными распределениями.

Сэмплирование из стандартных распределений

Существует несколько типов распределений, которые настолько широко используются, что считаются стандартными и обычно поддерживаются в популярных программных средах. MATLAB и Octave поддерживают многие типы вероятностных распределений, поэтому легко вычислить такие полезные показатели, как плотность вероятности, кумулятивная плотность, и выбрать случайные

значения из этих распределений. Полный список поддерживаемых распределений можно найти в документации MATLAB на странице «Поддерживаемые дистрибутивы», но наиболее популярные из них представлены в табл. 8. Как видно из таблицы, MATLAB может моделировать такие популярные распределения, как нормальное, экспоненциальное, пуассоновское. В таблице также представлены команды для расчета функции плотности вероятности, кумулятивной функции распределения и генерации случайных чисел.

Таблица 8

Примеры функций MATLAB для различных типов распределений

Distribution	PDF	CDF	RNG
Normal	normpdf	normcdf	norm
Uniform (continuous)	unifpdf	unifcdf	unifrnd
Beta	betapdf	betacdf	betarnd
Exponential	exppdf	expcdf	exprnd
Uniform (discrete)	unidpdf	unidcdf	unidrnd
Binomial	binopdf	binocdf	binornd
Multinomial	mnpdf	—	mnrnd
Poisson	poisspdf	poisscdf	poissrnd

Чтобы проиллюстрировать, как эти функции могут быть использованы, давайте визуализируем нормальное распределение, где $\mu = 100$ и $\sigma = 15$, которое могло бы, например, представлять собой изменение коэффициента IQ в определенной популяции. Код 5 показывает, как плотности вероятности и кумулятивной плотности могут отображаться вместе с гистограммой распределения. Как мы уже упоминали ранее, PDF и CDF широко используются для анализа вероятности. Результаты выполнения скрипта показаны на рис. 27.

Код 5. PDF, CDF и гистограмма для нормального распределения (выполнено в MATLAB/Octave)

```
mu = 100;% среднее значение
sigma = 15;% стандартное отклонение
xmin = 70;% минимальное значение x для pdf и cdf
```

```

xmax = 130;% максимальное значение x для pdf и cdf
n = 100;% количество точек на графиках pdf и cdf
k = 10000;% количество случайных ситуаций для гистограммы
x = linspace(xmin, xmax, n);% установить значения в промежутке
от xmin до xmax
p = normpdf(x, mu, sigma);% вычисление pdf
c = normcdf(x, mu, sigma);% вычисление cdf
subplot(1,3,1);
plot(x, p);
xlabel('x'); ylabel('pdf');
title('Probability Density Function');
% пауза
subplot(1,3,2);
plot(x, c);
xlabel('x'); ylabel('cdf');
title('Cumulative Density Function');
% получить k случайных чисел из N( $\mu$ ,  $\sigma$ ) — нормального распре-
деления
y = normrnd(mu, sigma, k, 1);
% пауза
subplot(1,3,3);
hist(y, 20);
xlabel('x'); ylabel('frequency');
title('Histogram of random values');

```

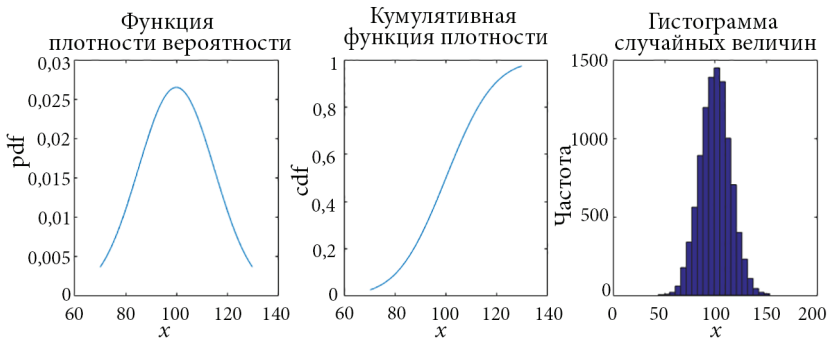


Рис. 27. Нормальное распределение для $\mu = 100$ и $\sigma = 15$

Аналогично рис. 28 демонстрирует в столбчатых диаграммах функции плотности и гистограмму для биномиального распределения, что часто встречается в ситуациях, когда исследователь подсчитывает количество успехов в общем числе испытаний. Мы рассмотрели несколько подобных ситуаций в предыдущей части. На рисунке показана ситуация, когда имеется 10 испытаний с вероятностью успеха 0,7.

Итак, как мы показали, MATLAB/Octave предоставляет ряд инструментов для анализа стандартных распределений, что позволяет легко использовать их в прикладных исследованиях в различных областях науки.

Сэмплирование из нестандартных распределений

В ситуациях моделирования часто возникает необходимость в использовании нестандартных распределений, которые могут быть или абсолютно новыми, или же являться комбинацией уже известных. Такие «нестандартные» распределения не поддерживаются программными пакетами. В этом случае вычислительные задачи часто опираются на сэмплирование уже известных распределений. Случайные значения, взятые из этих распределений, преобразуются, чтобы соответствовать целевому распределению. Некоторые из таких методов даже используются MATLAB для сэм-

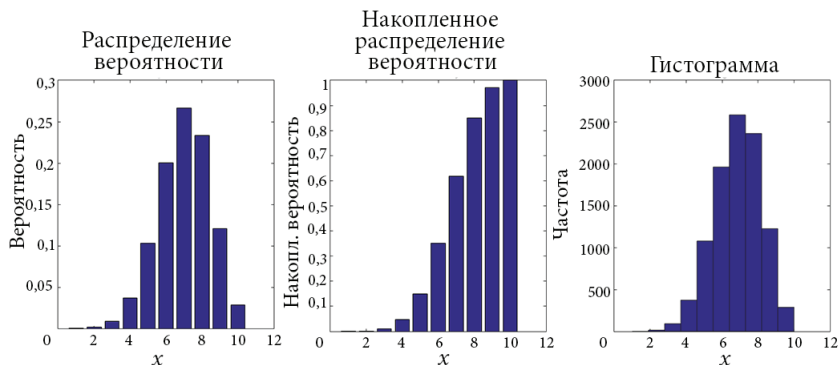


Рис. 28. Биномиальное распределение для $N = 10$ и $\theta = 0.7$

плирования стандартных распределений, таких как нормальное или экспоненциальное.

Сэмплирование обратных преобразований с дискретными переменными. Метод обратного преобразования — это метод сэмплирования для генерации случайных чисел из любого распределения, который основан на его обратной кумулятивной функции распределения. Данный метод выбирает равномерно распределенные случайные числа от 0 до 1 и преобразует их с помощью обратной кумулятивной функции распределения. Этот метод следует считать достаточно простым и практически универсальным, так как сэмплирование основано на преобразованных равномерных отклонениях. Например, MATLAB реализует некоторые из своих генераторов случайных чисел с помощью этой процедуры.

Приведем пример использования такого подхода. Самый простой способ проиллюстрировать этот подход — использовать дискретное распределение, где известна вероятность каждого отдельного исхода. В этом случае выборка обратных преобразований отображается как простой поиск в таблице. Мы используем данные о том, насколько хорошо люди могут «генерировать» равномерные случайные числа. В этой серии экспериментов людям предлагалось придумать большое количество случайных цифр, в свою очередь, исследователи подсчитывали частоту каждой случайной цифры. Очевидно, что люди далеко не всегда придумывают числа с универсальными распределениями. В табл. 9 приведены некоторые результаты одного из участков этих экспериментов. X представляет сгенерированную цифру; $\text{Pr}(X)$ и $\text{cdf}(X)$ представляют вероятность исхода и кумулятивные вероятности соответственно.

Таблица 9

Вероятность каждой из цифр, наблюдаемая
в эксперименте генерации случайных цифр человеком

X	$\text{Pr}(X)$	$\text{cdf}(x) (X)$
0	0,000	0,000
1	0,100	0,100
2	0,090	0,190

X	$\Pr(X)$	$\text{cdf}(x) (X)$
3	0,095	0,285
4	0.200	0,485
5	0,175	0,660
6	0,190	0,850
7	0,050	0,900
8	0,100	1,000
9	0,000	1,000

Как указано в табл. 9, цифры 0 и 9 не были сгенерированы вообще, некоторые цифры были довольно частыми (например, 4 и 6), а некоторые недопредставлены. Это ясно показывает, что человеческий мозг не очень хорошо подходит для получения равномерно распределенных случайных чисел.

Продemonстрируем алгоритм, который мог бы имитировать этот процесс, производя цифры с теми же вероятностями, что и в табл. 9. То есть представленный алгоритм будет генерировать 0 с вероятностью 0; 1 — с вероятностью 0,1; 2 — с вероятностью 0,09 и т. д. В приведенном ниже коде 6 показано, как это можно сделать и представить с помощью стандартных функций MATLAB (рис. 29).

Код 6. Моделирование выборки для случайных цифр

```
clear all; clc;
theta = [0.000;0.100;0.090;0.095;0.200;0.175;0.190;0.050;0.100;0.000]
K = 10000; y = rand(1, K); % k случайных чисел
Digitset = 0:9;
for i = 1:10
    Theta(i) = sum(theta(1:i)); % создать cdf
end
for i = 1:9
    Y(i) = sum(y > Theta(i) & y <= Theta(i+1))
end
counts = bar(Y); % гистограмма моделированных процессов
```

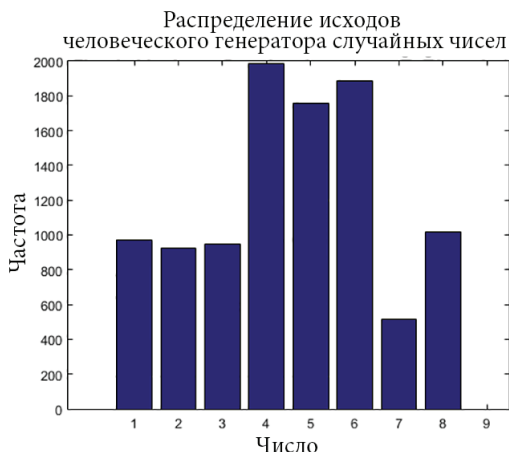


Рис. 29. Гистограмма генерированных результатов

```
xlim([-0.5 9.5]); xlabel('Digit'); ylabel('Frequency');
title('Distribution of simulated draws of human digit generator');
```

Гистограмма, очевидно, утверждает, что программа может в точности имитировать это распределение. Но давайте рассмотрим, как реализовать данный алгоритм, не используя встроенные функции, а используя метод обратного преобразования. Во-первых, кумулятивное распределение вероятности должно быть рассчитано, чтобы узнать, равна ли наблюдаемая вероятность результата определенному значению или меньше его. Поэтому нам необходимо рассчитать

$$\text{cdf}(X=x) = \Pr(X \leq x).$$

Для дискретных распределений это можно сделать с помощью простого суммирования. Результат операции показан в последней колонке табл. 9. Используя алгоритм обратного преобразования, мы должны выбрать равномерные случайные отклонения и сравнить каждое случайное значение с кумулятивными вероятностями в таблице. Первый результат, для которого случайное отклонение меньше (или равно) связанной кумулятивной вероятности, соответствует выборочному результату. Например, равномерное случайное отклонение $U = 0,8$ приводит к выборочному результату $X = 6$. Этот

алгоритм многократного сэмплирования однородных отклонений и сравнения их с кумулятивным распределением позволяет сформировать основу метода обратных преобразований для дискретных переменных. Обратите внимание, что в этой ситуации применяется обратная функция, так как выполняется обратный поиск в таблице.

Сэмплирование обратных преобразований с непрерывными переменными. Как уже упоминалось ранее, распределения также могут быть непрерывными, и для этого типа распределений также может быть применено сэмплирование с обратным преобразованием. В целом процесс преобразования аналогичен процессу преобразования дискретных переменных. То есть необходимо сгенерировать равномерные случайные отклонения и применить обратную функцию кумулятивного распределения к полученному значению. Пусть $\text{cdf}(X)$ представляет собой интегральную функцию плотности целевой переменной X и $\text{cdf}^{-1}(X)$ является обратной к функции, если инверсия может быть вычислена. Это может быть сделано путем многократного «вытягивания» случайной величины U из равномерного распределения между 0 и 1, затем принимая $X = \text{cdf}^{-1}(U)$.

Мы будем использовать следующий пример в качестве иллюстрации. Предположим, нам нужно выбрать случайные числа из экспоненциального распределения. Для $\lambda > 0$ cdf будет

$$\text{cdf}(x | \lambda) = 1 - \exp\left(-\frac{x}{\lambda}\right). \text{ Обратная этой функции будет } \text{cdf}^{-1}(u | \lambda) =$$

$= -\log(1 - u)\lambda$. Поэтому мы используем следующий пример алгоритма:

1. Сгенерировать $U \sim \text{Uniform}(0, 1)$.
2. Установить $X = -\log(1 - u)\lambda$.
3. Повторить.

Как можно видеть, процесс почти такой же, как и для дискретного распределения, которое было подробно рассмотрено ранее.

Метод отбраковки. Недостатком метода обратных преобразований является то, что он не может быть использован в любой ситуации, так как обратная функция не всегда может быть вычислена. В качестве альтернативы можно использовать сэмплирование отбраковкой (алгоритмом принятия — отклонения) или подходы Монте-Карло марковской цепи (МСМС). Здесь мы представим обзор

первого варианта. Основное преимущество отбраковки заключается в том, что все сэмплы, полученные в процессе отбора, могут быть использованы сразу же в качестве сэмплов из целевого распределения, без какого-либо дополнительного процесса «подготовки».

Приведем следующую иллюстрацию. Предположим, что мы генерируем точки равномерно в окружности с центром в $(0, 0)$ и $R = 1$. И так как может быть трудно не выходить из области круга, в этой ситуации может применяться сэмплирование отбраковкой. Сначала мы генерируем значения (x, y) из квадрата, описывающего интересующий нас круг, после чего отбрасываем любые значения, которые лежат за пределами круга. Другими словами, мы отбрасываем все значения, для которых $x^2 + y^2 > 1$. Этот процесс показан на рис. 30. Таким образом, мы использовали довольно простое и известное распределение в качестве основы для формирования выборки из более сложного распределения.

Таким образом, сэмплирование отбраковкой позволяет генерировать наблюдения из сложного для сэмплирования распределения, где вероятность конкретной выборки может быть оценена. Другими словами, предположим, что у нас есть определенное распределение $p(\theta)$, и прямое сэмплирование из этого распределения трудно, но плотность вероятности или ее значение для определенного значения x могут быть оценены. В таком случае исследователю

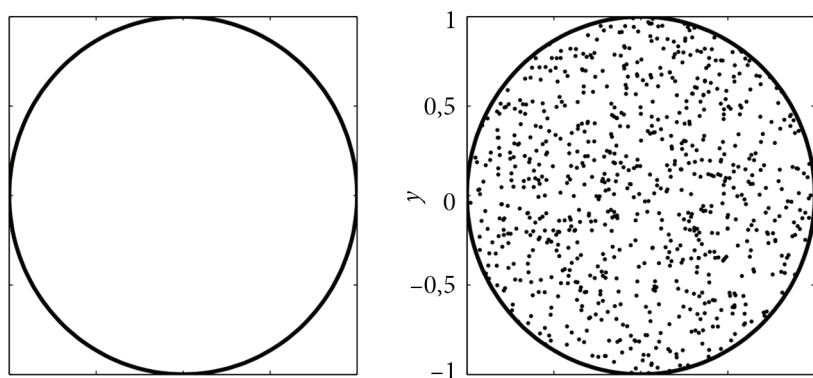


Рис. 30. Равномерная выборка из круга с использованием отбраковки [4, с. 13]

необходимо выбрать предлагаемое распределение. Распределение должно быть простым — $q(\theta)$, чтобы из него можно было непосредственно сформировать выборку. Затем исследователь оценивает вероятность взятых сэмплов при обоих распределениях, которые используются, и отбрасывает значения, так как они вряд ли будут находиться под целевым распределением относительно предложенного распределения.

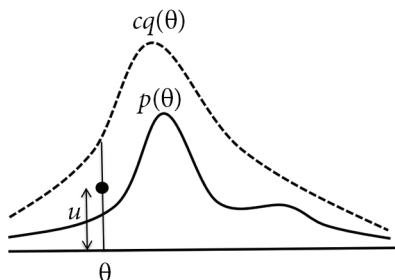


Рис. 31. Иллюстрация выборки методом отбраковки [4, с. 14]

Такая процедура показана на рис. 31. Сначала выберем константу c такую, что $cq(\theta) \geq p(\theta)$ для каждого возможного θ . Эта функция $cq(x)$, представляющая собой предложенную функцию $q(x)$, умноженную на константу c , всегда будет находиться «над» целевым распределением. Поиск соответствующей константы может быть нетривиальным, но обычно

выполняется с помощью некоторого исчисления. После выбора функции и константы мы генерируем значение u из распределения, которое лежит внутри $[0, cq(\theta)]$. Предложение отклоняется, если $u > p(\theta)$, и принимается в другом случае. Если предложение принято, выборочное значение θ берется из целевого распределения $p(\theta)$.

Подведем итог процедуры, указав шаги, которые предпринимаются в этой технике:

1. Выбрать распределение $q(\theta)$. Из этого распределения должно быть легко сформировать выборку.
2. Найти такую константу c , что $cq(\theta) \geq p(\theta)$ для всех θ .
3. Отобрать предложения θ из распределения предложений $q(\theta)$.
4. Выбрать равномерное отклонение u от интервала $[0, cq(\theta)]$.
5. Если $u > p(\theta)$, предложение отклоняется, в противном случае оно принимается.
6. Шаги 3, 4 и 5 повторяются до нужного количества значений; каждое принятое значение взято из $p(\theta)$.

Эффективность этой процедуры в значительной степени зависит от выбора предложенного распределения. Его следует выбирать так,

чтобы было принято как можно больше значений. Если распределение предложений сильно отличается от целевой функции, то многие сэмплы будут отклонены, что замедлит процедуру.

Представим код с использованием метода отбраковки (см. код 7). Предположим, мы хотим получить выборку из $Beta(2,1)$ для $0 \leq x \leq 1$. Код использует алгоритм, описанный выше, и в результате рисует гистограмму послецелевого распределения $f(x) = 2x$ для интервала $0 \leq x \leq 1$ (рис. 32).

Код 7. Техника сэмплирования методом отбраковки (MATLAB/Octave):

```
j = 1;
while i < 1000 % 1000 попыток
    y = rand;% взять Y из равномерного распределения [0, 1]
    u = rand;% взять U из равномерного распределения [0, 1]
    if u <= y% принимаем значение если U < 2Y/2
        x(j) = y;
        j = j + 1;
        i = i + 1;% попытка должна быть удачной
    end
end
hist(x);
```

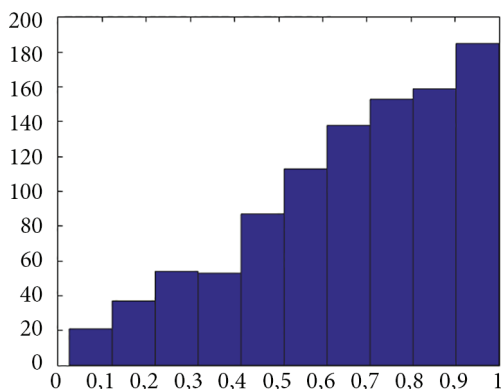


Рис. 32. Результирующая гистограмма $p(x) = 2x$

Значение c выбрано равным 2, так как в данном примере $\max(p(x)/q(x))=2$. Вероятно, для этого упражнения простое равномерное распределение предложений не является оптимальным, но оно используется, так как его легко реализовать в MATLAB. Очевидно, что увеличение количества генераций увеличит точность метода.

Этот метод также может быть расширен для работы с дискретными распределениями. В отличие от непрерывного случая, который мы рассмотрели (когда предложенное распределение $q(x)$ также должно быть дискретным).

Для высокоразмерных распределений может быть использован метод Монте-Карло марковской цепи, имеющий аналогичный подход. По существу, это вычислительный подход, который заменяет аналитическое интегрирование суммированием по выборкам, генерируемым из итерационных алгоритмов. В этом случае выборка по методу Монте-Карло позволяет оценить различные характеристики распределения, такие как среднее, дисперсия, эксцесс или любая другая интересующая статистика.

ГЛАВА 3. ЭВРИСТИЧЕСКИЕ МЕТОДЫ ОПТИМИЗАЦИИ

Другое применение методов Монте-Карло — решение сложных задач оптимизации. Для наиболее популярных методов оценки (например, оценка наименьших квадратов для линейных моделей) оптимизация проста и может быть решена с помощью простой алгебры. Однако иногда по ряду причин (например, мультиколлинеарность или просто большой размер модели) вычисление может быть проблематичным. Еще более сложной может быть оптимизация общих оценок максимального правдоподобия. В этом случае функции правдоподобия могут иметь несколько локальных оптимумов, что усложняет вычислительную процедуру.

Бывают и такие ситуации, когда традиционные методы не могут быть использованы для выбора модели из-за дискретности пространства поиска. Дискретная оптимизация обычно может быть решена путем перечисления потенциального решения, если это невозможно, то, вероятно, нет жизнеспособного детерминистического подхода к решению проблемы.

Как можно предположить из приведенных выше примеров, в эконометрическом анализе и моделировании стремительно растет число проблем, которые не могут быть решены с помощью традиционного подхода.

Существует три возможных подхода к указанной проблеме. Первый (он считается нецелесообразным, но иногда используется) — это игнорирование сложности и использование традиционных методов оптимизации с расчетом, что полученные результаты будут полезны в любом случае. Применение этого подхода может также объясняться непониманием реальной сложности предлагаемой модели. Конечно, это неправильно для решения такого рода проблем. Второй подход — упрощение модели до состояния, в котором можно использовать традиционную оптимизацию. В прошлом этот подход приходилось часто использовать из-за отсутствия вычислительной мощности. Он гораздо более адекватен, чем первый, но все же имеет свои недостатки, такие как жертвование возможными выгодами от сложных моделей и вероятность неправильной спецификации. Третий и наиболее предпочтительный подход — использование эвристической оптимизации. Этот вариант становится все более и более популярным в настоящее время, так как большинство компьютеров может обеспечить вычислительную мощность, необходимую для его эффективного использования.

Широко известно, что задача оптимизации является одной из центральных в эконометрике, будь то выбор наиболее подходящего типа модели или оценка ее параметров. Далее будет дан обзор эвристических методов оптимизации, в частности, методов Монте-Карло, и приведены несколько релевантных примеров.

Обзор эвристических методов

Термин «эвристика» обычно тесно связан с алгоритмами, имитирующими естественные процессы, такие как эволюция видов, процесс отжига твердых тел или самоорганизация колоний насекомых. Далее мы рассмотрим некоторые из этих алгоритмов. Более общее определение термина основывается на описании свойств алгоритма. Во-первых, эвристический алгоритм должен обеспечивать высокое качество аппроксимации глобального оптимума. Во-вторых, хорошо функционирующий эвристический алгоритм должен быть устойчив к изменениям характеристик задачи, т. е. иметь широкое применение для всего класса задач, а не только для одного случая. Кроме того, должна быть возможность настройки алгоритма путем изменения параметров или ограничений. В-третьих, что связано с предыдущим, эвристическая оптимизация должна быть применима для многих аналогичных проблем, в том числе и для новых. В эвристическом алгоритме не должно быть субъективных элементов.

Как видно из приведенных характеристик, одной из сильных сторон эвристики является отсутствие необходимости в строго фиксированном наборе допущений о задаче оптимизации для корректной работы метода. В большинстве случаев достаточно даже просто оценить целевую функцию для некоторого элемента пространства поиска. Нет необходимости предполагать какие-то глобальные свойства функции или вычислять производные. Недостатком этого типа алгоритмов оптимизации является то, что они дают не точные решения, а скорее их стохастические аппроксимации. Но тем не менее, если нет возможности использовать традиционные методы, эвристика может дать удовлетворительные результаты или оценки.

Сейчас, к сожалению, практическую реализацию эвристической оптимизации в эконометрике нельзя назвать устоявшейся по ряду причин. Первая, как мы уже упоминали ранее, — недостаток знаний или понимания сложности задачи оптимизации приводит к использованию стандартных методов вместо эвристических. Это ограничение весьма специфично и, к сожалению, его нелегко преодолеть из-за отсутствия инструментов для выявления неэффективности классических методов. Например, сравнение стандартного подхода

и эвристики может указывать на неадекватность классического инструмента только в том случае, если качество эвристического решения намного лучше, но не наоборот. Хорошее «классическое» решение не обязательно указывает на то, что подходит стандартный подход к оптимизации.

Вторая причина — отсутствие хорошо документированных реализаций эвристических подходов. Увеличение количества различных эвристических методов и гибридов может сбить с толку даже опытного исследователя. Решением этого вопроса легче заниматься, чем первым, и его можно было бы решить по мере появления все более доступных ориентиров для эвристических методов. Поскольку эвристические модели оптимизации обычно довольно просты в программировании, многие успешные реализации можно легко сравнить и проанализировать. Будем надеяться, что примеры и классификация наиболее распространенных эвристических методов, приведенные в этой книге, помогут преодолеть читателям эту проблему.

Третья причина — рассмотрение стохастической природы результатов эвристической оптимизации может создать трудности для исследователей. Что касается этого вопроса, то он тесно связан с предыдущим, и, вероятно, все большее число публикаций, касающихся теоретических и практических аспектов эвристической оптимизации, могло бы искоренить эту проблему.

Эвристические методы оптимизации, как правило, являются вычислительными, поэтому они развиваются параллельно развитию электронных вычислительных устройств. Первые вклады в эвристические методы оптимизации появились еще в 1950-х гг., когда предлагалось использовать алгоритмы для решения проблемы коммивояжера. Основные технологии в этой области были внедрены в конце 1980–1990-х гг., но лишь недавно получили широкое распространение, так как персональные компьютеры достигли производительности, необходимой для эффективного использования эвристических методов. За последние десятилетия количество эвристических методов значительно выросло и продолжает расти. Далее мы представим общие понятия эвристической оптимизации и предложим возможную классификацию некоторых существующих

вариантов. Следует отметить, что на сегодняшний день нет общепринятой «наилучшей» классификации эвристических алгоритмов оптимизации, однако представленная классификация может помочь выявить общие характеристики методов и сопоставить их.

Эвристические методы оптимизации (также называемые методами аппроксимации) обычно делятся на два класса: конструктивные методы и методы локального поиска. Методы первого класса, как правило, строят решение через последовательность локально оптимальных вариантов. Методы второго класса используют только информацию о решениях в окрестности текущего решения и, таким образом, очень похожи на алгоритм восхождения к вершине, где выбор соседнего решения локально максимизирует критерий. Этот класс методов аппроксимации долгое время не считался полезным и только недавно стал популярным. Приведем обобщенный алгоритм метода локального поиска минимизации заданной функции $f(x)$.

1. Сгенерировать начальное решение x^c .
2. Повторять шаги 3–4 до тех пор, пока не будут выполнены определенные критерии остановки (обычно определенное количество итераций).
3. Выбрать $x^n \in N(x^c)$, соседа текущего решения x^c .
4. Если $f(x^n) < f(x^c)$, то набор $x^c = x^n$. Поскольку мы минимизируем целевую функцию, чем меньше она становится, тем лучше.

Как можно заметить, этот алгоритм аналогичен алгоритму восхождения к вершине, где используется информация о градиенте для выбора соседа, но в данном случае x^n выбирается по случайному механизму. Механизм и критерии приемлемости определяют способ поиска решений алгоритмом.

Этот широкий класс методов можно разделить на два меньших класса: траекторные методы для одного решения и популяционные методы, которые производят набор решений. Первый класс включает метод порогового принятия (threshold accepting) и поиск табу (tabu search), второй — генетические алгоритмы (genetic algorithms), методы дифференциальной эволюции (differential evolution) и колонии муравьев (ant colony). Каждый метод имеет свои правила выбора соседа и принятия решения и в большинстве случаев допускает «подъемы», чтобы избежать локальных минимумов. Давайте теперь

рассмотрим эти методы, начиная с методов, которые считаются траекторными методами.

Правильное определение соседствующего решения имеет большое значение для этих методов и обычно зависит от рассматриваемой проблемы, однако найти это определение может быть сложно.

Алгоритм имитации отжига (Simulated Annealing), как хороший представитель траекторных методов (ТМ), использует вероятностный критерий приемлемости и основан на подобии комбинаторной оптимизации и процесса отжига твердых тел. Как и в классическом методе локального поиска, улучшение текущего решения всегда принимается. Более того, движения «в гору» также могут быть приняты, но с некоторой вероятностью, которая обычно реализуется с помощью ГСЧ с равномерным распределением. Эта вероятность зависит от разности значений функции и параметра T (называемого температурой применительно к процессу отжига), который постепенно уменьшается с каждым «витком» процесса. Поэтому в процессе работы вероятность принятия «менее оптимального» решения неуклонно снижается. Критерий остановки определяется либо общим числом итераций, либо числом последовательных итераций без улучшения текущего решения x^c . Следующий алгоритм представляет, как метод работает шаг за шагом.

1. Сгенерируем начальное решение x^c ; инициализируем количество испытаний R и температуру T .

2. Повторим шаги 3–7 R раз.

3. Будем повторять шаги 4–6 до тех пор, пока критерии останова не будут выполнены.

4. Вычислим $x^n \in N(x^c)$, соседнее для текущего решения x^c .

5. Вычислим $\Delta = f(x^n) - f(x^c)$ и сгенерируем равномерную случайную величину u .

6. Если $\Delta < 0$ или $e^{-\Delta/T} > u$, то установим $x^c = x^n$.

7. Уменьшим T .

Метод порогового принятия (ТА) является, по существу, детерминированным вариантом метода SA с последовательностью температур T , замененной последовательностью детерминированных порогов T . Более подробный обзор алгоритма мы дадим ниже.

Метод поиска табу (TS) — метод, предназначенный для анализа дискретных пространств поиска с конечным множеством соседних решений. Метод реализует выбор соседнего решения, что позволяет избежать цикличности, другими словами, одно и то же решение не может быть выбрано дважды. Это достигается за счет использования кратковременной памяти, списка табу, который содержит последние выбранные решения. Шаги алгоритма следующие:

1. Сгенерировать начальное решение x^c ; инициализировать список табу $Tabu = \emptyset$.

2. Повторять шаги 3–5 до тех пор, пока критерии остановки не будут выполнены.

3. Вычислить
$$V = \frac{\{x \mid x \in N(x^c)\}}{T}.$$

4. Выбрать $x'' = \min(V)$.

5. Установить $x^c = x''$ и $Tabu = Tabu \cup x''$.

В этом случае множество V может быть построено на основе соседних решений или, в качестве альтернативы, можно использовать более сложный алгоритм. Обычно $Tabu$ обновляется путем простого удаления старых значений и добавления новых.

Популяционные методы. Траекторные методы, как упоминалось ранее, работают только над одним решением, в то время как популяционные методы могут работать одновременно с набором решений (называемым популяцией). Учитывая это, популяционные методы могут быть гораздо более эффективными для изучения всего пространства поиска и, потенциально, обеспечить более точные результаты за счет сложности и более высоких требований к вычислительной мощности. Рассмотрим несколько алгоритмов этого метода.

Генетический алгоритм (GA) получил огромную популярность. Этот тип алгоритмов поиска предназначен для имитации эволюционного процесса видов, размножающихся половым путем. Новые кандидаты для решения генерируются с помощью механизма скрещивания, при котором объединяется часть характеристик каждого родителя, а затем применяется случайная мутация. Если новый «индивид», называемый ребенком, унаследует хорошие характе-

ристики от своих родителей, то у него будет большая вероятность выжить. Соответствие ребенка и родительской популяции оценивается функцией выживания p , которая может быть сформирована из нескольких источников: последний созданный индивид P' , $P' \cup \{fittest\ from\ P'\}$, только сильнее из P' или сильнее из $P' \cup P'$. Вот пример алгоритма, представляющего это.

1. Сгенерировать совокупность начальных решений P .
2. Повторять шаги 3–9 до тех пор, пока не будут выполнены условия остановки.
3. Выбрать $P' \subset P$ (брачный набор), инициализировать $P'' = \emptyset$ (набор потомков).
4. Повторить шаги 5–8 для требуемого числа итераций.
5. Выбрать индивидов x^a и x^b случайным образом от P' .
6. Применить перекрестное соотнесение x^a и x^b , чтобы произвести x^{child} .
7. Случайным образом мутировать полученный x^{child} .
8. $P'' = P'' \cup x^{child}$.
9. $P = survive(P', P'')$.

Другими словами, алгоритм в целом принимает множество решений на шаге 3, а затем строит множество соседних решений с выбранным критерием.

Другой популяционный метод называется *Ant colonies* — АС (также известный как колония муравьев) и имитирует то, как муравьи ищут пищу и возвращаются в колонию. Во-первых, исследование окрестностей носит случайный характер. Как только муравей находит пищу, он начинает транспортировать пищу обратно в гнездо, оставляя феромонный след на пути, который приведет других муравьев к пище. Интенсивность следа указывает не только на качество и количество пищи, но и на расстояние между гнездами и пищей. Чем короче расстояние, тем больше муравьев будет путешествовать, т. е. феромоны будут сильнее. Такое поведение способно оптимизировать работу муравьев, так как они следуют по следам своих сородичей. Этот естественный механизм оптимизации был успешно адаптирован в эвристике. В этом случае область поиска муравья соответствует дискретному набору, из которого выбираются элементы решений, количество пищи связывается с целевой

функцией, а след феромона моделируется с помощью адаптивной памяти. Алгоритм работает с каждой итерацией, улучшая решение до тех пор, пока не будет достигнута цель оптимизации.

Третий алгоритм этого класса называется *дифференциальной эволюцией* (DE). Алгоритм использует эволюционный подход и обновляет совокупность векторов решений путем сложения, вычитания и пересечения их, а затем выбирает наиболее подходящие решения среди исходных и обновленных наборов. Начальное заполнение n_p случайно выбранных решений может быть представлено матрицей $P^{(0)}$ размера $d \times n_p$, где d — размерность области функции (рис. 33).

Алгоритм строит n_G поколений популяции. Для получения нового поколения требуется несколько шагов:

1. Для каждого решения $P_i^{(0)}$, $i = 1, \dots, n_p$ (выделенный столбец на рис. 33), алгоритм создает два промежуточных решения $P_i^{(v)}$ и $P_i^{(u)}$ из трех случайно выбранных столбцов $P_{r_1}^{(0)}$, $P_{r_2}^{(0)}$ и $P_{r_3}^{(0)}$.

2. I -е решение $P_i^{(1)}$ нового поколения собирается из промежуточных решений и начального.

3. Затем эти шаги повторяются до тех пор, пока проблема оптимизации не будет решена.

Процесс также контролируется дополнительными параметрами, влияющими на скорость усадки при исследовании области решений, или вероятностью пересечений. Эти параметры зависят от рассматриваемой проблемы, поэтому выбираются в зависимости от популяции или какой-либо другой специфики.

Стохастика является преобладающей темой в методах оптимизации, которая в то же время отличается от методов стохастического случайного сэмплирования. В случае, когда многократные применения метода оптимизации не дают идентичных результатов,

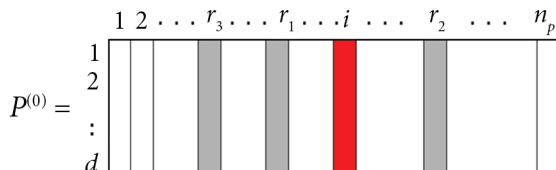


Рис. 33. Начальная популяция решений в DE-алгоритме [5, с. 14]

приходится иметь дело с этим типом стохастичности. Как можно было догадаться, для методов оптимизации, рассмотренных ранее, многие факторы могут определять случайность, например, генерация начального решения, выбор решения «кандидата» или процесс принятия. С одной стороны, судя по практической стороне, дополнительная случайность крайне нежелательна для эвристики. Однако, с другой стороны, из-за ограничений классических подходов они могут обеспечить детерминированные результаты, которые в некоторых случаях далеки от оптимальных. В подобных ситуациях очевидно, что стохастическая аппроксимация истинного оптимума намного лучше, чем плохой детерминированный результат. Поэтому стохастичность результатов эвристики следует пересмотреть.

Рассмотрим $\psi^{L,r}$ как результат запуска R некоторых эвристических алгоритмов оптимизации H для заданной целевой функции f . Пусть I — количество ресурсов, затраченных на один запуск алгоритма, и $f(\psi^{L,r})$ представляют собой полученные значения целевой функции, которая может рассматриваться как случайная выборка из распределения $D_I^H(\mu_I, \delta_I, \dots)$ с определенным конечным математическим ожиданием и дисперсией. Стоит отметить, что хотя для большинства эвристик свойства этого распределения неизвестны, некоторые общие характеристики могут быть получены. Во-первых, для задач минимизации распределение будет подвергнуто цензуре на глобальном минимуме целевой функции. Во-вторых, с увеличением объема вычислительных ресурсов I распределение, как правило, смещается влево и становится менее рассеянным.

При многократном применении эвристической оптимизации к одной и той же задаче становится возможным вычисление эмпирического распределения целевой функции. Эта информация, в свою очередь, позволяет оценить свойства D_I^H и его эмпирические характеристики. В частности, более низкие значения интересны как оценки наилучших решений задачи минимизации. Теория экстремальных значений может быть использована для получения оценок неизвестного глобального. Наконец, повторение алгоритма для различных объемов вычислительных ресурсов позволяет оценить эмпирическую скорость сходимости.

На рис. 34 представлены результаты применения упомянутого ранее метода порогового принятия к задаче выбора лаговой структуры в VAR-моделях. Верхний левый график показывает эмпирические функции распределения целевой функции (байесовский информационный критерий) для различных значений I . По мере роста I распределения перемещаются влево и становятся менее рассеянными. На трех других графиках показаны оценки плотности ядра распределения $f(y^I)$ для различного числа повторений.

Эти графики подтверждают выводы эмпирической функции распределения. Кроме того, они могут подсказывать информацию о конкретных свойствах целевой функции. Например, фигура на нижнем правом графике, вероятно, указывает на два локальных минимума с большой областью притяжения, причем первый может быть глобальным минимумом.

Сходимость эвристики

Как мы уже упоминали, оптимизационная эвристика не дает детерминированного результата, поэтому важно описать сходимость этих методов в рамках нашего обзора. Интуитивно понятно, что увеличение количества вычислительных ресурсов (например, увеличение общего числа запусков) должно положительно влиять на точность результата, если алгоритм настроен правильно.

Подобная логика может применяться и для других эвристических методов. Эта работа в целом подтверждает теоретическое предположение о том, что с увеличением I результаты алгоритма становятся более точными. Давайте рассмотрим это более подробно.

Во-первых, результаты, очевидно, зависят от нескольких предположений. Если быть точным, то пространство поиска должно быть связано, т. е. из любой заданной начальной точки можно добраться до любой другой точки пространства поиска, выполнив локальные шаги поиска, а целевая функция должна удовлетворять определенному критерию. Во-вторых, авторы доказывают, что существуют подходящие параметры для пороговой реализации, допускающей такое, что глобальный оптимум целевой функции может быть ап-

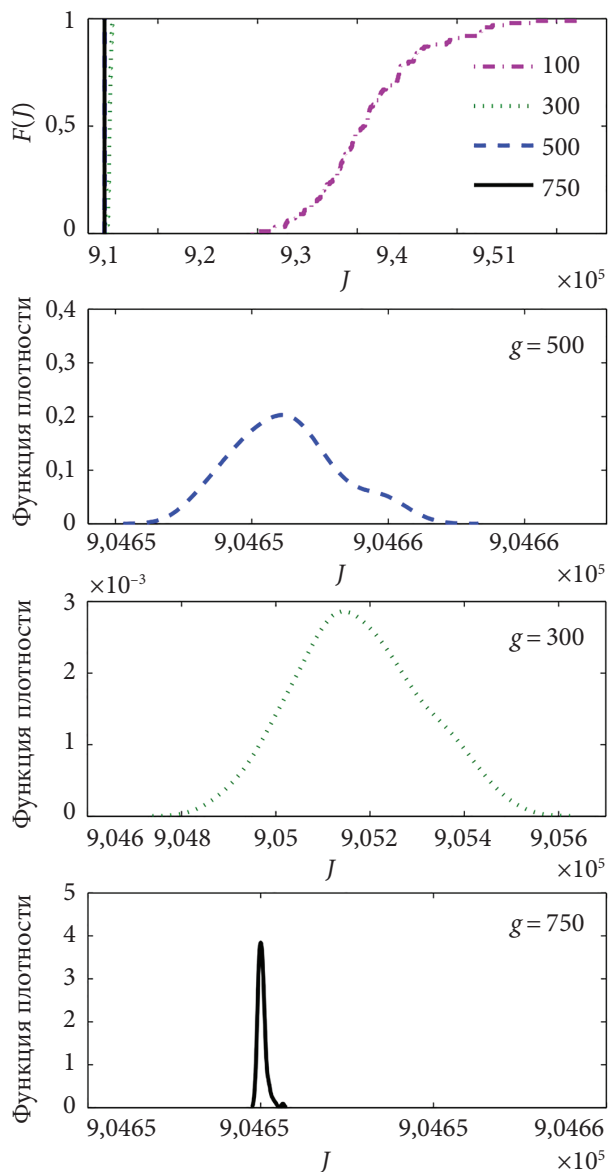


Рис. 34. Эмпирическое распределение F для разных значений J ($n = 1000$) [6, с. 12]

проксимирован с произвольной точностью с любой заданной вероятностью, близкой к одной, путем увеличения числа итераций. Однако на практике найти подходящие значения параметров может быть довольно сложно.

Может быть представлен следующий результат сходимости: при любом $\delta > 0$ и $\varepsilon > 0$ существует число итераций $I(\delta, \varepsilon)$ такое, что для любого заданного повторения r

$$P\left(\left|f\left(\psi^{I,r}\right)-f_{\min }\right|<\varepsilon\right)>1-\delta,$$

где f_{\min} обозначает глобальный минимум целевой функции. Есть несколько замечаний, которые следует упомянуть в рамках этого утверждения. Главным является то, что результат не дает никакой информации о скорости сходимости, что крайне важно, так как нет возможности жертвовать бесконечные ресурсы на алгоритм. Кроме того, для любого конечного количества I существует ошибка в аппроксимации оценок. Поэтому необходимы дальнейшие исследования в этой области.

Руководство по использованию оптимизационных эвристик.

Поскольку мы рассмотрели ключевые основы эвристических методов оптимизации, мы теперь дадим несколько полезных рекомендаций по использованию этих методов в практических исследованиях. Сначала объясним, когда используются эти методы. Вообще, для ситуации, когда задача оптимизации сложна в вычислительном смысле, например, из-за нескольких локальных минимумов, в качестве ориентира для «традиционного» подхода следует использовать эвристическую оптимизацию. Если в этом случае эвристическая оптимизация даст лучшие результаты, чем классический подход, то это должно быть поводом для применения эвристики в качестве основного метода оптимизации. Возможно, эвристику можно было бы даже использовать в качестве стандартного бенчмарка, так как многие программные пакеты ее поддерживают.

Также при использовании эвристики для оптимизации исследователь должен правильно выбрать тип применяемого алгоритма. Во-первых, выбор может быть основан на типе пространства поиска и целевой функции, поскольку некоторые эвристики, такие

как дифференциальная эволюция, не подходят для дискретного пространства поиска или проблем с параметрами шума в функции. Во-вторых, исследователи могли бы использовать предыдущий опыт применения эвристики в некоторых областях, где проблемы имеют схожие характеристики. Кроме того, следует начать с более общих подходов, а затем, при необходимости, использовать более конкретные эвристические методы.

Каждый из перечисленных выше алгоритмов оптимизации имеет свой собственный набор настраиваемых параметров, которые должны контролироваться исследователем для повышения эффективности алгоритма. Например, в траекторных методах окрестность решений должна быть определена тщательно, а соседнее решение должно быть легко генерируемым. В генетическом алгоритме происходит обмен информацией между особями во время скрещивания с мутациями, таким образом, следует контролировать информацию, передаваемую от родителей, и ее разнообразие.

Проиллюстрируем реализацию эвристической оптимизации методом порогового принятия. Для этого рассмотрим задачу минимизации на подмножестве Ω из R^k . Реализация этого алгоритма включает в себя указание количества перезапусков $n_{restarts}$, количества раундов n_{rounds} и количества шагов n_{steps} , а затем определение целевой функции, окрестности и пороговой последовательности. В этом случае число итераций на перезапуск I определяется числом раундов, умноженным на число шагов.

Оптимизация с точки зрения вычислений имеет решающее значение для эвристики, поэтому при построении алгоритма исследователь всегда должен стремиться к достижению наиболее быстрого расчета. Для повышения производительности метода порогового принятия целевая функция должна обновляться локально, т. е. при изменении функция x не должна вычисляться с нуля, а просто обновляться. Такой подход также полезен в популяционных алгоритмах. К сожалению, на практике локальное обновление функций может быть проблематичным.

Далее следует рассмотреть наличие ограничений, и в этом случае Ω является подпространством R^k , и создание начальных и соседних решений может быть затруднено, особенно если Ω не связано.

Поэтому R^k следует использовать в качестве пространства поиска с добавлением определенного штрафа к целевой функции, если x не принадлежит пространству Ω . Обычно на практике этот штраф изначально устанавливается небольшим, чтобы позволить исследовать все пространство, а затем увеличивается с каждым запуском, чтобы увеличить точность решения. Штрафы могут быть представлены в абсолютном или относительном выражении, причем второй вариант допускает более широкое использование.

Теперь перейдем к определению окрестности. Как правило, в случае ТА для вещественных переменных определение может быть дано с помощью ε -сфер:

$$N(x^c) = \{x^n \mid x^n \in R^k \text{ and } \|x^n - x^c\| < \varepsilon\},$$

где N обозначает окрестности, а $\|\dots\|$ — евклидову метрику, которая в случае дискретного пространства поиска может быть заменена метрикой Хэмминга для получения лучших результатов.

К сожалению, теоретический анализ метода порогового принятия не дает рекомендаций по выбору пороговой последовательности. В некоторых случаях малых задач оптимальная пороговая последовательность не уменьшается монотонно. К счастью, существуют два подхода, которые, как представляется, обеспечивают надежные пороговые последовательности для использования в области эконометрики. Первый и самый простой — использовать последовательность, линейно убывающую к нулю за выбранное количество периодов, в этом случае настраивать нужно только первое пороговое значение. Второй подход заключается в использовании управляемой данными (data driven) генерации последовательности, которая опирается на саму структуру функции. Рассмотрим этот вариант подробнее.

Для случая конечного пространства поиска актуальны только пороговые значения, соответствующие разности значений целевой функции для пары соседей. Чтобы быть применимым к реальным пространствам поиска, алгоритм использует случайную выборку из распределения таких локальных разностей. Эта процедура также может быть успешно применена к случаям бесконечных и непре-

рывных пространств поиска. Следующий алгоритм дает общее представление о том, как работает этот подход:

1. Случайным образом выбрать $x^c \in \Omega$.
2. Повторить шаги 3–4 выбранное количество раз n_{delta} .
3. Рассчитать $x^n \in N(x^c)$ и $\Delta_i = |f(x^c) - f(x^n)|$.
4. Принять $x^c = x^n$.
5. Рассчитать эмпирическое распределение F по Δ_i .
6. Рассчитать последовательность порогов $\tau_r = F^{-1}\left(\frac{n_{rounds} - r}{n_{rounds}}\right)$, $r = 1, \dots, n_{rounds}$.

Полученное эмпирическое распределение Δ показано на рис. 35.

Вместо того, чтобы рассматривать локальные изменения целевой функции Δ вдоль пути через пространство поиска, как описано в алгоритме, исследователь может рассмотреть несколько перезапусков, чтобы сгенерировать различные траектории меньшей длины или — в пределе — генерировать все x^c случайным образом. Следует отметить, что все три метода должны давать одинаковую аппроксимацию распределению локальных изменений целевой функции с бесконечным количеством попыток.

Далее мы рассмотрим, как можно контролировать алгоритм. Чтобы понять, как работает алгоритм, можно построить график значений функций, принятых в последующих раундах. На рис. 36 показан такой график (мы использовали порог, определенный ранее). Как видно, с увеличением успешных раундов подъемы происходят реже. В последнем раунде вообще нет подъемов.

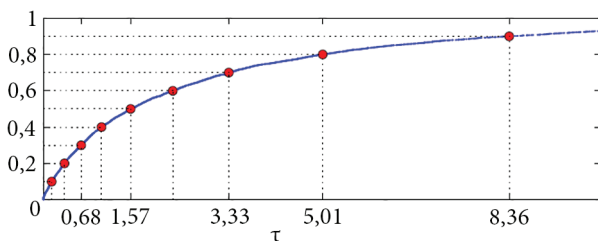


Рис. 35. Эмпирическое распределение последовательности Δ :
 τ — показатель числа раундов [5, с. 31]

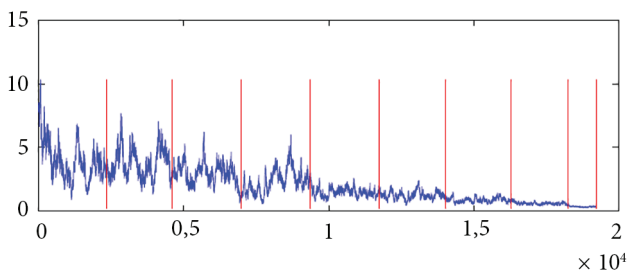


Рис. 36. Значения функций для принятых шагов локального поиска, разделенные на раунды [5, с. 31]

Два последних графика могут предоставить исследователю полезную информацию о том, насколько хорошо заданы параметры алгоритма. Например, неправильная форма кумулятивного распределения Δ должна быть сигналом плохого локального поведения целевой функции. По графику принятых значений также можно определить, правильно ли выбрано количество шагов и раундов. Обычно количество шагов должно намного превышать количество раундов, как правило, предполагается, что алгоритм по умолчанию имеет не менее 10 раундов, и увеличение этого числа повышает точность. Очевидно, что выбор этих параметров зависит от решаемой проблемы, поэтому если оценка целевой функции является дорогостоящей, они будут выбраны насколько возможно малыми.

Хотя, как мы упоминали ранее, оценки обычно выигрывают от увеличения числа итераций I , и было бы разумно разделить имеющиеся ресурсы на несколько перезапусков. Во многих случаях оптимально установить количество перезапусков $n_{restarts}$ от 5 до 20, а не тратить все имеющиеся ресурсы за один раз. Если перезапуски выполняются в распределенной вычислительной среде, оптимальное распределение вычислительных ресурсов должно рассматриваться по-другому. Опять же вопрос в том, как выбрать $n_{restarts}$ и I , чтобы минимизировать время выполнения для заданного качества результата $f_{sol} = \min_{r=1, \dots, n_{restarts}} f(\psi^{I,r}) \leq c$.

На рис. 37 показано эмпирическое распределение результатов $f(\psi^I)$ эвристической оптимизации для увеличения значения I . Авторы используют случайную величину Z Бернулли в решении

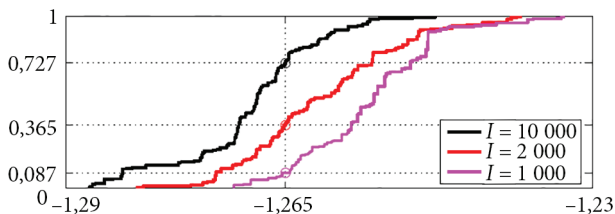


Рис. 37. Эмпирическое распределение результатов $f(\psi^1)$ [5, с. 33]

$f(\psi^1)$ в одном исполнении, где $Z = 1$, если $f(\psi^1) < c$, и $Z = 0$ — иначе. Для $I = 10\,000$ и $c = -1,265$ они получили $p = P(z = 1) = 0,727$, что соответствует процентилу от $f(\psi^1) = -1,265$ в эмпирическом распределении. Затем авторы рассмотрели случайную величину x , которая подсчитывает количество решений, удовлетворяющих $f(\psi^1) < c$ из всех решений. Если $x \sim B(n_{\text{restarts}}, p)$ является биномиальной случайной величиной, то вероятность того, что решение по крайней мере настолько хорошо, как $f_{\text{sol}} = \min_{r=1, \dots, n_{\text{restarts}}} f(\psi^{1,r}) \leq c$, является следующей: $\pi = 1 - P(x = 0)$. Далее можно вычислить количество перезапусков, необходимых для получения требуемого качества c с заданной вероятностью π . Для $x \sim B(n, p)$ мы имеем $P(x = 0) = (1 - p)^n$, и число перезапусков является наименьшим положительным целым числом N , подходящим для $\pi \leq (1 - p)^n$.

В результате авторы пришли к выводу, что $I = 1000$ и $n_{\text{restarts}} = 51$ будет самым лучшим выбором в распределенной вычислительной среде, по меньшей мере, с 51 узлом, в то время как $I = 2000$ и $n_{\text{restarts}} = 11$ — это лучший выбор, если перезапуски должны быть выполнены последовательно. Кроме того, они упоминают, что алгоритм ТА с перезапусками ведет себя устойчиво в данном классе проблем.

Далее рассмотрим несколько возможных применений эвристической оптимизации для дискретных и непрерывных пространств поиска.

Приложение к дискретным пространствам поиска

Рассмотрим ряд примеров. В первом примере эвристика ТА используется как иллюстрация выбора модели в моделях VAR.

Рассмотрим p -мерный вектор AR-процесса X_t , который задается

$$X_t = \sum_{k=1}^K \Gamma_k X_{t-k} + \varepsilon_t,$$

где матрицы Γ_k обеспечивают коэффициенты авторегрессии и $\varepsilon_t \sim N(0, \Sigma) iid$. В этом случае все компонентные временные ряды являются стационарными.

Предположим, что максимальная длина лага K_0 может быть получена либо из экономического бэкграунда, либо из определенного эмпирического правила, основанного на некотором количестве наблюдений. Таким образом, модель отбора проблемы для данной реализации процесса включает в себя определение K и ненулевых элементов Γ_k в пространстве поиска, описываемого $\Omega = \{0, 1\}^{p^2 \times K_0}$, где 0 указывает на параметр ограничения, равного нулю, а 1 — на свободно оцениваемый параметр. Можно также отметить, что для этой задачи могут быть рассмотрены различные целевые функции. Можно использовать байесовский информационный критерий (информационный критерий Шварца) $BIC = \ln |\hat{\Sigma}| + l \ln T/T$ (где $\hat{\Sigma}$ обозначает определитель ковариационной матрицы подогнанных невязок, а l — число ненулевых параметров), но можно взять любой другой информационный критерий (например, информационный критерий Акаике). Это можно было бы реализовать, просто заменив целевую функцию.

Предполагается, что Ω конечно, что позволяет решить задачу простым перечислением всех элементов множества и выбором лаговой структуры, соответствующей наименьшему значению выбранного информационного критерия. Однако этот метод не применим к реальной задаче, так как даже скромные значения p^2 и K_0 выдают большое количество лаговых структур, что делает перечисление практически невозможным. Другой способ — рассмотреть только небольшое подпространство Ω , которое можно перебрать. Однако в таком случае нет никакой гарантии, что оптимальная структура

лага будет лежать в этом подмножестве. Поэтому использование эвристики целесообразно хотя бы в качестве ориентира или отправной точки.

Как упоминалось в предыдущем параграфе, окрестность для эвристики можно установить с помощью ε -сфер, определяемых расстоянием Хэмминга. В этом случае, как правило, ε должно быть выбрано достаточно большим, чтобы алгоритм не «застрял» в плохих локальных минимумах, и достаточно маленьким, чтобы привести к управляемому локальному поиску. Для рассматриваемой задачи с 5000 итераций на перезапуск расстояние Хэмминга, равное 4, оказывается наиболее адекватным выбором, так как стандартное отклонение принятых значений является наименьшим (табл. 10). Пороговая последовательность строится по описанным ранее алгоритмам с 30 раундами по 100 шагов каждый.

Таблица 10

Стандартное отклонение локальных изменений и принятых решений

distance	$\sigma_{\Delta N}$	$\sigma_{accepted}$
2	0,1706	0,9569
4	0,2578	0,5961
6	0,2916	0,7932

Рассмотрим задачу с $p = 3$ и $K = 3$ (для простоты). Эта задача демонстрирует различные лаговые структуры для трех уравнений и нескольких неопределенностей, так что стандартный подход не может привести к правильному выбору модели. Процедура отбора основана на смоделированных данных этого процесса с $T = 100$ и $K_0 = 5$ для целей оптимизации.

Далее приведем несколько примеров реализаций процесса с 1000 перезапусков, но с другого числа итераций I . Для $I = 500$ глобальное лучшее значение встречается всего 4 раза, для $I = 100$ – 18 раз, для $I = 5\,000$ – 217 раз, а для $I = 10\,000$ – 332 раза. При этом с увеличением I среднее относительное отклонение от лучшего решения уменьшается с 22 % в первой попытке до 1,3 % в последней. При небольшом количестве наблюдений за реализацией процесса не следует ожидать, что модель с наименьшим значением информационного

критерия, найденная эвристикой, соответствует истинному процессу генерации данных. Эвристика просто обеспечивает лучшее или, по крайней мере, хорошее приближение для данной реализации. Поэтому для получения точных результатов следует использовать большой набор различных структур моделей и реализаций для каждой из них.

Приложение к непрерывным пространствам поиска

Далее приведем сравнение реализаций порогового принятия и дифференциальной эволюции в случае задачи непрерывной оптимизации, рассмотрим модель линейной регрессии:

$$y_i = \begin{bmatrix} x_{i1} & \dots & x_{ip} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix} + \varepsilon_i, \quad i=1, \dots, n,$$

где матрица θ обозначает коэффициенты и $\varepsilon \sim N(0, \sigma^2 I)$ — вектор ошибок. В этом случае наименьший процент загрязнения в данных, который может привести к произвольному смещению оценки, является точкой перелома оценки. Для нормального OLS перелома оценка — 0 %, но есть некоторые другие оценки, для которых это значение 50 %. Фактически оценки с высокой точкой перелома не могут быть решены так просто, как МНК, потому что их целевые функции невыпуклы и имеют несколько локальных минимумов. За последние два десятилетия для решения этой проблемы был предложен ряд алгоритмов. Эти методы, по сути, используют свертки. Сравнительно недавно алгоритмы были пересмотрены и оптимизированы для более быстрого выполнения. Полученные алгоритмы являются сложными специальными процедурами. Мы демонстрируем, что стандартные методы эвристической оптимизации могут быть использованы для решения таких задач.

Рассмотрим пример наименьшей медианы квадратов (LMS):

$$\hat{\theta}_{LMS} = \underset{\theta}{\operatorname{argmin}} Q_{LMS}(\theta),$$

где $Q_{LMS}(\theta) = \text{med}(r_1^2, \dots, r_n^2)$ — это медиана квадратов остатков r_i^2 . Кроме того, используются наименьшие урезанные квадраты (least trimmed squares) и S-оценка.

Рассмотрим модель линейной регрессии, упомянутую ранее, где 90 % переменных являются независимыми и одинаково распределенными случайными величинами с нормальными распределениями, поэтому $\theta_i = 0$, $i = 1, \dots, p$. Остальные 10 % состоят из выбросов, соответствующих отклонениям $\theta_2 = M/100$ и $\theta_j = 0$ для $j \neq 2$.

Данные генерируются для $n = 100$, $p = 10$ и $M = 190$, а затем вычисляются оценки LMS. Следует отметить, что эти оценки должны существенно отличаться от нуля. Рассмотрим рис. 38, иллюстрирующий невыпуклую форму целевой функции. Левая часть представляет данные, которые были сгенерированы с помощью искусственной модели, представленной выше. Локальные минимумы для этой функции соответствуют $\theta_2 = 1,9$. В правой части представлены реальные данные, связанные с биомедицинской проблемой.

Далее мы приступим к возможной реализации алгоритма ТА для задачи минимизации с непрерывной целевой функцией. Как ранее упоминалось, $Q_{LMS}(\theta) = \text{med}(r_1^2, \dots, r_n^2)$, где $r_i = (y_i - X_i, \theta)^2$, $i = 1, \dots, n$. Для генерации решения θ^n используется следующий алгоритм:

1. Установить $\theta^n = \theta^c$.
2. Выбрать равномерно $i \in \{1, \dots, p\}$.
3. Сгенерировать равномерную случайную величину $u \in [-h, h]$.

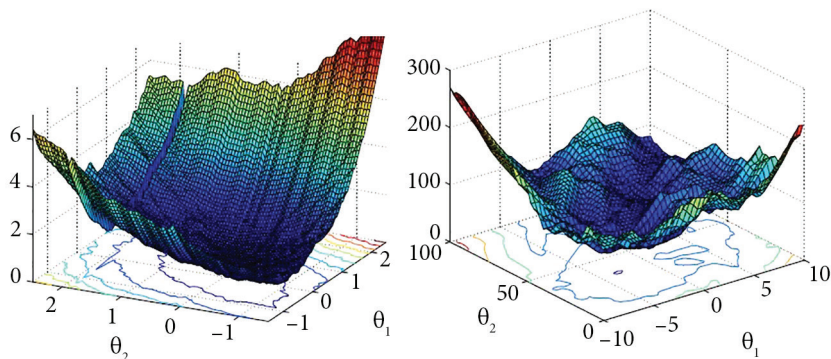


Рис. 38. Форма функции $Q_{LMS}(\theta)$ для LMS [5, с. 40]

4. Рассчитать $\theta_i^c = \theta_i^c + u(1 + |\theta_i^c|)$.

В этой симуляции h установлен на значении 0,40 и алгоритмы выполняют 25 тыс. итераций, разделенных на 10 раундов по 2500 шагов. Результаты можно увидеть на рис. 35 и 36.

Другим способом решения проблемы является использование алгоритма DE. При этом целевая функция остается неизменной. Чтобы сделать результаты сопоставимыми, мы задаем $n_p = 100$ и $n_G = 250$, что приводит к тому же числу итераций, что и в первой ситуации. Другими соответствующими параметрами являются: $CR = 0,8$ и $F = 0,75$, $\theta_i \in [-3,3]$ для $i = 1, \dots, p$. Эмпирическое распределение значения целевой функции для TA и DE показано на рис. 39.

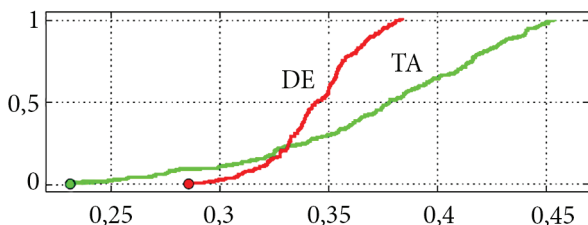


Рис. 39. Эмпирическое распределение значения целевой функции двухста решений для TA и DE [5, с. 42]

В рассматриваемой задаче все решения показывают наличие выбросов, т. е. ни одно из оцениваемых θ_2 не равно 1,9. Мы могли бы также заключить, судя по графику, что для этого случая дифференциальная эволюция дает несколько лучшие решения, чем TA. Очень небольшая дисперсия решений также указывает на то, что для получения с высокой вероятностью решения, идентифицирующего выбросы, требуется лишь небольшое число перезапусков, около 10.

ГЛАВА 4. ВВЕДЕНИЕ В АГЕНТО-БАЗИРОВАННОЕ МОДЕЛИРОВАНИЕ

Агенто-базированные модели (ABM) или, как их иначе называют, индивидуальные модели (IBMs) являются типом вычислительных моделей, которые используются при моделировании поведения автономных агентов для оценки того, как ведет себя система, состоящая из нескольких агентов. Эта группа моделей достаточно новая, но доказала свою эффективность в экономической теории и многих социальных науках. Случайность, необходимая для моделирования, достигается с помощью методов Монте-Карло.

По сравнению с подходами на основе переменных, использующими структурные уравнения, или системными подходами, использующими дифференциальные уравнения, агенто-базированное моделирование способно имитировать индивидуальную неоднородность, которая явно представляет правила принятия решений агентов и помещает агентов в географическом или другом типе пространства, что дает практически бесконечное количество возможных применений этого метода.

В этой главе рассматривается понятие агенто-базированных моделей, то, как они работают, чем отличаются от других типов моделей, и дается несколько возможных путей реализации агенто-ориентированных моделей в исследовании, также приводится несколько полезных примеров моделей на основе агентов.

Обзор агенто-базированных моделей

Агенто-базированное моделирование (Agent-Based Modelling) — это, по сути, вычислительный метод, позволяющий создавать, анализировать и экспериментировать со сложными моделями, состоящими из автономных агентов, взаимодействующих со средой и друг с другом. Этот подход эффективно сочетает теорию игр, вычислительную социологию, мультиагентные системы и эволюционное программирование.

Агнто-базированное моделирование представляет собой вычислительный подход, таким образом, включает в себя создание моделей и алгоритмов, которые могут быть выполнены с использованием компьютерных ресурсов. В данном случае слово «моделирование» используется в общем смысле, как создание упрощенного представления реальных явлений или объекта, несущего в себе «отличительные черты». В эконометрике, например, регрессионное уравнение с зависимой переменной и регрессорами служит моделью отношений между переменными. Даже словесное описание можно рассматривать как модель: например, описывая, как прямые иностранные инвестиции влияют на инновационную деятельность в той или иной стране, мы создаем модель реальных процессов, достаточно простую.

Вычислительные модели обычно представлены в виде компьютерных программ, состоящих из нескольких входных параметров, которые, следуя какому-либо алгоритму, приводят к выходным значениям, представляющим моделируемый процесс. В качестве примера можно привести предположения о том, как реклама или социальные отношения влияют на вкусы потребителей. Для их тестирования необходимо создать компьютерную программу, в которой есть набор агентов (или частных лиц), которые делают выбор покупки в соответствии со своими предпочтениями. В этом случае на каждого агента будет влиять выбор его «друзей» (и наоборот) и реклама как некая «внешняя» сила. А поскольку традиционные методы представляются слишком сложными для использования в моделях с большим количеством элементов, то такие процессы можно моделировать, применяя агентный подход.

Существует несколько преимуществ вычислительного моделирования. Во-первых, оно использует точный компьютерный язык, возвращая таким образом точные численные результаты, в то время как теории и модели на естественном языке иногда могут быть неопределенными или формально оцененными. Во-вторых, компьютерные программы хороши для моделирования процессов, поскольку среды программирования обычно предоставляют для этого различные инструменты. В-третьих, как мы уже неоднократно отмечали ранее, вычислительная мощность машин быстро растет,

что позволяет моделировать более сложные, а иногда и непредсказуемые процессы с участием многих событий или агентов.

Далее следует учитывать, что АВМ позволяют экспериментировать с моделями. Это важно главным образом потому, что в социальных науках, в отличие от физики или химии, например, экспериментировать очень сложно или чрезвычайно дорого, главным образом потому, что классический метод «лечения — контроля» не может быть применен в социальной сфере. В физико-химических исследованиях эксперимент состоит из применения определенной обработки к изолированной системе, а затем контроля результатов, сравнения этой системы с другой, которая не получила никакого лечения. Очевидно, то же самое нельзя было применить к социальным исследованиям в связи со сложностью социальных систем и по ряду этических причин. Однако использование агентного моделирования в исследованиях нивелирует трудности изоляции человеческой системы и этические проблемы экспериментирования, поэтому такие эксперименты предпочтительнее проводить на виртуальных системах. Это можно также назвать одним из основных преимуществ АВМ. Желаемый эксперимент можно настроить и провести практически неограниченное количество раз, применив к системе ряд экзогенных параметров и даже позволив появиться случайным вариациям.

Эту идею нельзя назвать новой. Например, архитекторы или инженеры тестируют модели зданий или самолетов в аэродинамической трубе, так как гораздо дешевле и проще использовать модель, чем реальный объект. Другая причина для экспериментов с моделями заключается в том, что иногда это может быть единственным способом получения результатов, как, например, в квантовой физике, поскольку частицы почти невозможно наблюдать непосредственно. Рассмотрение поведения модели аналитически обычно предпочтительнее, так как оно предоставляет информацию о том, как модель будет вести себя с учетом диапазона входных данных. Для этих случаев необходимо экспериментировать с различными входными данными модели, чтобы получить информацию о ее поведении. Та же логика применяется и к социальным экспериментам.

Очевидно, что для достижения достоверных результатов такого рода экспериментов исследователь должен убедиться, что используемая модель ведет себя так же, как и реальная система, что иногда может быть сложным. Таким образом, экспериментирование на моделях не является универсальным решением.

Другая основная концепция этой главы — модель. Модели имеют долгую историю использования в социальных исследованиях. Моделирование социальных процессов происходило еще до использования компьютеров, но получило широкое распространение, когда статистические модели стали использоваться как инструмент анализа больших объемов данных в экономике, социологии и демографии. Как уже упоминалось ранее, модели предназначены для моделирования некоторых реальных явлений, которые называются целью модели. Два основных их преимущества заключаются в том, что они лаконично выражают отношения между характеристиками цели и позволяет обнаружить ее новые характеристики при исследовании модели.

Одна из самых ранних моделей социальных наук — гидравлическая модель экономики Филлипса, также называемая аналоговым компьютером денежного национального дохода (MONIAC), в которой вода, проходящая через систему труб и сосудов, используется для моделирования обращения денег в экономике. Модель также позволяет изменять параметры, например, процентную ставку, изменяя скорость течения воды по трубам, чтобы увидеть, как это влияет на всю систему. Рассмотрим несколько наиболее распространенных типов моделей.

Масштабные модели — просто уменьшенная версия реального объекта, но помимо уменьшения его размера они также упрощают его. Например, масштабная модель здания будет иметь ту же форму, что и реальное строение, но не будет отражать его систему водоснабжения или электроснабжения. Карту города также можно назвать масштабной моделью, так как она, очевидно, намного меньше, чем сам город, и содержит в себе информацию только о положении зданий, но не о их высоте или других параметрах. При изучении этого типа моделей следует помнить, что результаты модели должны быть масштабированы до размеров исследуемого

объекта, и учитывать возможность того, что несмоделированные элементы могут повлиять на результат.

Другим типом моделей являются *идеализированные модели*. В этом случае некоторые характеристики цели преувеличены для упрощения модели. Можно привести пример идеализированной модели фондового рынка, которая предполагает, что информация мгновенно перетекает от одного трейдера к другому, и идеализированной модели дорожного движения, которая предполагает, что водители всегда знают идеальный маршрут. Основная идея идеализации в моделировании состоит в том, чтобы убрать из модели сложные факторы, оказывающее незначительное влияние на модель, тем самым упростив ее и получив жизнеспособные результаты.

Модели, основанные на проведении аналогии между уже изученным явлением и объектом, называются *аналогическими моделями*. Одним из наиболее известных примеров является бильярдная модель атомов. Примеры из социальных наук — это компьютерная модель разума и модель мусорного бака организаций. Эти модели обычно очень полезны, так как явления, используемые для аналогии, обычно достаточно известны и могут обеспечить хорошие оценки, если аналогия адекватна.

Четвертая категория — *математические модели* или *модели, основанные на уравнениях*. Эти модели наиболее часто используются в экономике и количественной социологии. Обычно они определяют отношения между переменными, но в отличие от предыдущих моделей они не используют аналогии или сходства между моделью и целью. Вместо этого они используют математические функции, чтобы описать эти отношения. Качество такого рода моделей определяется тем, насколько хорошо их данные описываются построенным уравнением. В качестве примера такого рода моделей можно упомянуть производственную функцию Кобба — Дугласа. Это широко используемая математическая модель, отражающая связь выпуска и затраченных ресурсов:

$$Y = AL^{\alpha}K^{\beta},$$

где Y обозначает выпуск, L — трудозатраты, K — капитальные затраты, а A , α и β — константы, определяемые текущей технологией.

Форма этого уравнения была получена с использованием данных, а не путем теоретических предположений. Математические модели чрезвычайно полезны при анализе того, как связаны переменные или явления, но обычно они не дают четкой информации о том, почему эти явления связаны. Поэтому другие типы моделей лучше, если процесс или механизм связи является основной проблемой.

Как мы уже упоминали, агенто-базированные модели воссоздают ситуации, когда агенты взаимодействуют в определенной среде. Агенты используются в социальных науках для представления отдельных лиц, организаций, фирм или даже стран. Обычно это части программы (а иногда и отдельные программы), которые реагируют на свою вычислительную среду как на модель реальной социальной среды. Особенность АВМ состоит в том, что агенты способны взаимодействовать, передавая друг другу информацию посредством диалога, наблюдения или каким-либо другим способом. Эта особенность отличает агенто-базированное моделирование от других подходов.

Давайте подробнее рассмотрим, что такое окружающая среда в АВМ. Это может быть нейтральный виртуальный мир, который мало или совсем не влияет на агентов или, наоборот, оказывает на них сильное влияние. Окружающая среда может представлять собой географическое пространство, например, в моделях, касающихся международных отношений, окружающая среда может представлять особенности страны или ее регионов. Модели, в которых окружающая среда представляет собой географическое пространство, называются пространственно-явными. В этом случае, как правило, у агентов есть координаты, которые отражают их местоположение в виртуальной среде. В других типах моделей среда также представляет собой определенное пространство, но не географическое, а, например, «пространство знаний». Агенты также могут быть связаны друг с другом в сети, где отношения одного агента с другими показывают его положение в сети.

Существует несколько основных критериев «хорошей» АВМ. Например, К. Френкен [7] представляет пять таких критериев, давайте рассмотрим их. Во-первых, АВМ обычно используют «стилизированные факты», такие как широко распространенные эмпи-

рические нормы, а не точные эмпирические данные, что может вызвать проблемы излишней идентификации. Для преодоления этих проблем исследователь может использовать дополнительные критерии выбора наиболее подходящей модели, а также комплементарные модели построения, повторяющие несколько упрощенных фактов в одной модели, для широкого диапазона или вероятного диапазона значений параметров. Поэтому если модель повторяет большое количество упрощенных фактов, то она считается хорошей по критерию «тиражирования стилизованных фактов» (*ROSF*). Кроме того, такой критерий способен оценить, насколько хорошо существующие модели могут учитывать новые стилизованные факты, и тем самым оценить надежность модели.

Другой популярный критерий, который обычно используется для моделей эволюции, называется *ROAS* (реализм предположений). Это означает, что рассматриваемая модель не должна противоречить известным поведенческим фактам или фактам окружения. В неоклассических моделях используется противоположная этому инструменталистская философия, согласно которой главным критерием качества моделей является способность делать точные предсказания.

Третьим и, возможно, самым основным является критерий качества — *KISS*, который расшифровывается как «Keep it Simple, Stupid». Это обычно понимается как рекомендация начать моделирование с простой модели, которая может быть легко понятна и давать понятные результаты, и только затем развивать модель в определенном направлении. В этом смысле такая рекомендация оспаривается недавно разработанным принципом *KIDS* (Keep it Descriptive Stupid), предложенным в работе Б. Эдмонса и С. Мосса [8]. Согласно этому принципу моделирование должно проводиться самым простым из возможных способов, с использованием самого широкого диапазона фактических данных. И упрощение используется только тогда, когда оно оправдано данными или самой моделью. Модели *KISS* изначально просты, что потенциально может привести к нежелательному упрощению. В качестве альтернативы принцип *KISS* может применяться постфактум: для проверки того, является ли модель самой простой моделью, которая может дать желаемые

результаты, и выполняются ли критерии *ROSF* и *ROAS*. В настоящее время существует ряд мнений, поддерживающих идею о том, что простые модели более подходят для оценки реальных данных, чем сложные модели, которые обычно мотивированы желанием более точно имитировать реальность.

Последний критерий, известный как «возьмите предыдущую модель и добавьте что-нибудь», сокращенно *TAPAS*, утверждает, что инкрементальный подход обычно является наиболее эффективным. Этот критерий предназначен для эвристического моделирования, но также может быть использован при оптимизации модели. Основной идеей данного критерия является использование существующих моделей и предложений в качестве основы для максимальной кумулятивности знаний и минимизации времени на разработку новой модели, причем такой подход делает новые модели более понятными, в отличие от моделей, построенных с нуля. Это не свидетельствует о том, что нет места для экспериментальной инновационной модели в АВМ, а скорее означает то, что если можно воспользоваться уже существующей моделью, то ей стоит воспользоваться, сократив время на построение и упростив понимание новой модели. С учетом последних двух принципов очевидно, что значительная часть моделирования на основе агентов должна быть посвящена построению простых базовых моделей, которые затем можно было бы легко модифицировать для более конкретного использования.

В заключение этой главы рассмотрим отличительные особенности агентного моделирования, которые изложены и объяснены в работе М. Гилберта [9]. Каждый из них иллюстрирует, чем АВМ отличается от других типов моделей.

Онтологическое соответствие. Агентные модели обеспечивают прямое соответствие между реальными действующими агентами и моделируемыми агентами, что упрощает проектирование и интерпретацию модели по сравнению с моделями на основе уравнений. Агентная модель производственной фирмы может включать агентов, представляющих поставщиков, сотрудников (которые могут быть разделены по ролям на различные группы), клиентов и любых других возможных субъектов. В модели может использоваться только один агент для каждого класса субъектов или же

отдельный агент для каждого основного субъекта в среде, если это важно для результатов.

Представление окружающей среды. Еще одной особенностью АВМ является возможность непосредственного представления реальной среды в модели, независимо от того, включает ли она физические характеристики (например, расстояние между агентами или в регионе), воздействие окружающих агентов или влияние других факторов, таких как изобилие ресурсов или скученность. Эта особенность имеет решающее значение для построения надежных моделей промышленных районов, которые будут упомянуты в следующем разделе. Еще один пример применения такого метода — это моделирование движения туристов в горах Сьерра-Невада в Калифорнии для оценки влияния различных стратегий управления на сохранение этой области дикой природы. В этой модели моделировалось перемещение агентов на ландшафте, представляющем топологию области.

Гетерогенные агенты. Общеизвестно, что традиционные подходы обычно касаются объединенных субъектов. Иными словами, все рассмотренные субъекты схожи по своим важнейшим чертам. Например, экономика обычно имеет дело с «типичной фирмой» или «типичным экономическим агентом», чьи решения являются рациональными. Эти агенты могут отличаться в своих предпочтениях в определенных моделях, но следовать тому же поведению. Даже если существуют различные группы субъектов с их собственным поведением, их число невелико главным образом потому, что если агенты не однородны, то получить осмысляемые аналитические выводы, как правило, невозможно. АВМ преодолевают это ограничение, позволяя каждому агенту действовать в соответствии со своими предпочтениями и правилами. В данном случае примером могут служить модели цепи поставок, где у каждого бизнеса в цепи своя стратегия.

Взаимодействие агентов. Одним из важнейших преимуществ АВМ является возможность взаимодействия между агентами. Другими словами, в этом типе моделей агенты могут передавать информацию друг другу, это может быть так же просто, как передача сообщений между двумя соседними агентами, или может быть гораздо сложнее и включать в себя различные интерпретации одного

сообщения. Хорошей иллюстрацией этой особенности являются модели динамики мнений.

Ограниченная рациональность. Обычно в традиционных экономических моделях индивиды, действующие на рынке, считаются рациональными. Это можно было бы расценивать как слабость такого рода моделей, поскольку многие исследователи предполагают, что люди «гиперрациональны» в своем стремлении следовать сложным или даже бесконечным цепочкам логики, чтобы выбрать оптимальный курс действий, что, очевидно, нереально. Справедливости ради следует отметить, что существуют и модели с агентами, которые действуют иррационально или хаотично и не будут максимизировать свое благосостояние, но количество таких моделей относительно невелико. В качестве альтернативы преобладающей полной рациональности возможно моделировать индивидов с ограниченной рациональностью, другими словами, с ограниченными возможностями оптимизации из-за ограниченных когнитивных способностей. АВМ не только позволяют ограничить рациональность агентов, но и потенциально позволяют агентам принимать решения такого же уровня сложности, как и решения людей. В качестве примера можно привести несколько моделей фондовых рынков.

Обучение. Еще одной полезной особенностью АВМ является то, что они позволяют моделировать процесс обучения агента как на индивидуальном, так и на социальном уровне. Такие модели, как инновационные сети (которые будут упомянуты в следующей главе), позволяют фирмам на рынке научиться производить более прибыльный и желаемый продукт. Более того, сектора в целом могут обучиться эффективно работать, а продукция фирм в секторе может производить совместимый набор, как если бы одна фирма покупала компоненты для собственного производства у других фирм. Существует три наиболее известных подхода к моделированию процесса обучения:

1. Индивидуальное обучение, которое предполагает, что каждый агент учится на собственном опыте отдельно от других.
2. Социальное обучение, в котором одни агенты подражают другим или учат их. Это приводит к обмену индивидуальным опытом среди всего населения.

3. Эволюционное обучение. Этот тип обучения предполагает, что некоторые агенты уходят с рынка и заменяются более конкурентоспособными агентами, которые учат других агентов, как успешно работать.

Естественно, что при необходимости эти виды обучения могут комбинироваться. Например, вышеупомянутые инновационные сети используют следующий механизм обучения: отдельные инновационные фирмы учатся производить более качественную продукцию, а поскольку плохо функционирующие фирмы становятся банкротами и заменяются более эффективными стартапами, сектор в целом может научиться улучшать свои показатели.

Ознакомившись с концепцией агенто-базированного моделирования и рассмотрев его основные отличительные особенности, представим несколько полезных примеров использования АВМ для конкретизации и уточнения приведенных выше определений, а также возможные пути реализации такого вида моделирования в реальных исследованиях.

Примеры использования АВМ

АВМ могут использоваться в ряде общественных наук. В качестве примера рассмотрим отдельные проблемы, при которых можно было бы эффективно использовать агентные модели.

Городские модели. Предложенная в 1971 г. первая модель города была направлена на объяснение расовой сегрегации в американских городах. Эта модель была очень абстрактной, что оказало влияние на следующие работы, касающиеся сохранения сегрегации не только в городских центрах по всему миру. Модель состоит из квадратной сетки ячеек, которая представляет собой городскую область, где агенты-домохозяйства распределены случайным образом. Для простоты эти два вида агентов обычно называются «красными» и «зелеными». Каждая ячейка сетки может иметь только одно домашнее хозяйство, поэтому некоторые ячейки остаются пустыми. Алгоритм выполняет следующие шаги:

1. На каждом этапе каждая семья собирает информацию о своих восьми соседях.

2. Домохозяйство проверяет, превышает ли доля соседних домохозяйств «с другим цветом кожи» определенный уровень «толерантности».

3. Если доля больше, чем уровень допуска (например, есть 6 «красных» домохозяйств, окружающих «зеленое» домохозяйство с 0,5 значением допуска), домохозяйство считает себя «несчастливым» и перемещается в пустую ячейку.

4. «Счастливое» домохозяйство остается в своей позиции до следующего шага.

В связи с тем, что домохозяйства каждый раз переезжают, именованная доля для каждого домохозяйства постоянно меняется, заставляя его пересматривать свои решения на каждом этапе, что приводит к новому переезду. Рисунок 40 иллюстрирует, как первоначально случайное распределение (левая панель) домохозяйств реорганизуется в сплошные группы красного и зеленого цвета (правая панель).

Кластеризация происходит даже тогда, когда домохозяйства терпимо относятся к большинству соседей с отличным цветом кожи, что было заявлено как показатель того, что даже высокий уровень расовой толерантности позволяет формировать сильно сегрегированные модели, характерные для городов США.

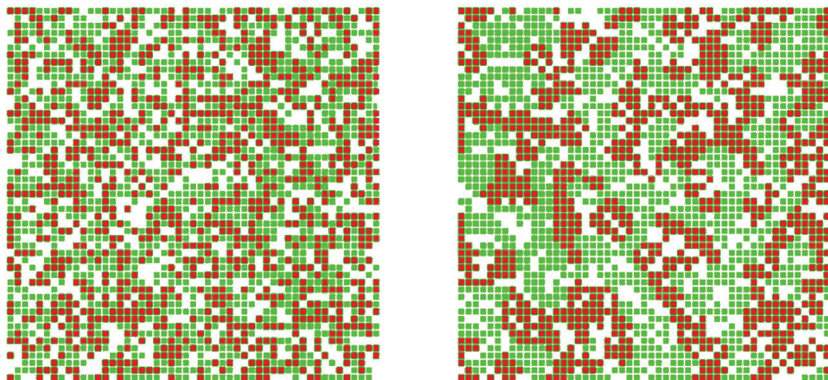


Рис. 40. Пример модели Шеллинга (порог допуска = 0,3) [9, с. 7]

Есть несколько основных причин, почему эта модель была так влиятельна. Во-первых, результат легко предсказать, зная правило движения каждого агента. Во-вторых, сама модель проста и зависит только от одного параметра — значения допуска. В-третьих, результирующее поведение устойчиво к различным значениям порога допуска и начального размещения. Те же результаты получены для широкого диапазона значений параметров, правил движения и рассматриваемых соседей. К примеру, агент мог принимать случайные решения о движении, рассматривать полезность незанятой ячейки или учитывать цену ячеек и другие сценарии. В-четвертых, модель предполагает тестирование на эмпирических данных. Однако на практике она оказалась довольно проблематичной из-за трудности получения достоверных данных и измерения «несчастья».

Таким образом, эта модель была популярна из-за простоты состава и несложного получения результатов. В настоящее время данная модель совершенствуется путем добавления дополнительных параметров и модифицируется, чтобы быть более подходящей для реального мира.

Модели динамики мнений. Вторая группа моделей важна для понимания того, как политические или любые другие мнения развиваются среди населения. Есть ряд работ, рассматривающих эту проблему, мы представим одну из них. Один из самых интересных вопросов в этой области исследований, вероятно, вопрос о том, как маргинальные или даже экстремальные мнения могут стать нормой для большей части населения. Есть много исторических примеров, иллюстрирующих это явление. Самый простой пример связан с манерой одеваться: мы видим, что «экстремистский» стиль одежды с течением времени стал социально приемлемым и распространенным, например, ношение мини-юбок. К сожалению, в истории есть и гораздо более драматичные примеры. Можно вспомнить процессы в Германии 1930-х гг. или на Ближнем Востоке в последние десятилетия, где радикальному меньшинству удалось заставить значительную долю населения принять их политические убеждения. Есть также много фактов, свидетельствующих о сильной биполяризации населения, когда оно все целиком становилось экстремистским, например, Европа периода «холодной войны».

В модели, предложенной Деффонтом и другими исследователями, агенты имеют мнение, которое является действительным числом от -1 до 1 , и в определенной степени сомневаются в своем мнении. Эти параметры вместе образуют сегмент мнения агента, который представлен в виде линии, учитывающей мнение агентов и протяженной по степени неопределенности. Модель состоит из следующих шагов:

1. Каждый агент начинает с мнения, сформированного случайным образом из равномерного распределения и с общего уровня неопределенности, за исключением небольшого количества экстремистов, которые достаточно уверены в своем мнении.

2. Агенты случайно встречаются друг с другом.

3. При встрече один агент (i) может влиять на мнение другого (j), если их сегменты мнения пересекаются. В противном случае они считаются слишком разными, чтобы влиять друг на друга.

4. Если один агент оказывает влияние на другого, то его мнение изменяется пропорционально разнице между мнениями, умноженной на количество перекрытий, деленное на неопределенность агента i , минус один. Это указывает на то, что люди, сильные в своих убеждениях, влияют на других больше, чем неопределенные агенты.

После выполнения этих шагов экстремизм распространяется до тех пор, пока все агенты не присоединяются к этому течению, или каждый агент становится «традиционалистом». Если симуляция проходит без экстремистов, то все население занимает среднюю позицию. Эта модель демонстрирует то, как небольшая группа экстремистов с сильными убеждениями может оказать катастрофическое влияние на общество в целом.

Модели потребительского поведения. Эта группа моделей, как правило, используются предприятиями, чтобы лучше понять, что именно мотивирует клиентов покупать продукцию. Социальные факторы, такие как влияние друзей и семьи, реклама, мода, имеют важное значение, но зачастую их трудно оценить. Это можно сделать с помощью агентных моделей. Одним из первых применений такого подхода было исследование ситуации, при которой один из конкурирующих продуктов на рынке достигает такого доминирования, что отдельным потребителям трудно переключиться на конкуриру-

ющие продукты (эту ситуацию обычно называют «блокировкой»). В качестве примеров такой ситуации можно назвать клавиатуры с QWERTY-раскладкой (доминирующие над другими клавиатурами), VHS, консоли Gameboy или операционные системы Windows. Исследователи в основном сосредоточились на поведенческом процессе, который приводит к блокировке, таким образом, они построили модель с «consumats» (агентами), имеющими определенные правила принятия решений, которые адаптированы к общим поведенческим теориям социального сравнения и имитации.

Другим примером такого рода моделей является исследование рынка подержанных автомобилей и других рынков с изменчивостью и неопределенностью качества. Исследование демонстрирует, как различное качество различных изделий может влиять на доверие агентов и рынок в целом. Исследователи представили два типа агентов в модели: покупатели и продавцы. Рынок содержит конечное количество продуктов, с которыми покупатели и продавцы работают, выполняя одну сделку за раунд, после чего рынок очищается. Продавцы действуют исходя из минимальной цены продукта, которую они готовы принять, в то время как покупатели предлагают оплату на основе ожидаемого качества продукта. Покупатели делают свои выводы о качестве исходя из индивидуального и социального опыта. Социальная сеть создается посредством случайного соединения пар агентов с определенной скоростью, которая контролируется изначально заданным параметром.

Результаты исследования показали важность социальных связей для такого рода рынков. Если покупателям приходится полагаться только на индивидуальный опыт, доверие потребителей падает и рынок рушится, тогда как обмен опытом в социальных сетях позволяет поддерживать стабильный рынок. Кроме того, это показывает, как социальная информация может агрегировать опыт группы на более высокий уровень и, таким образом, уменьшить важность плохого опыта одного человека.

Модели промышленных сетей. Ранее экономическая теория не уделяла большого внимания устоявшимся связям между фирмами, но в последние десятилетия исследователи приходят к пониманию того, что сети чрезвычайно важны в промышленных

секторах. Есть известный пример, касающийся промышленных районов на севере Италии, таких как Прато. Основная характеристика промышленного района заключается в том, что в небольшом географическом регионе сосредоточено большое количество мелких фирм, которые производят один и тот же продукт и имеют прочные связи между собой. Данные связи по-разному представлены в ряде исследований, например, это могут быть отношения между поставщиком и потребителем, обмен технологиями производства, финансовые связи или даже дружеские отношения.

Другим популярным примером являются инновационные сети, которые можно наблюдать в таких наукоемких секторах, как компьютерные технологии. Эти сети могут несколько отличаться от стандартных промышленных районов, поскольку они не требуют географической близости фирм в секторе, но соединения одинаковы. Рассмотрим пример, когда в инновационных сетях агенты обладают определенным запасом знаний и опыта (так называемый «kenes»). Эти «kenes» используются для создания новых продуктов, которые затем продаются другим фирмам из близкого окружения. Однако продукт может быть произведен только в том случае, если компоненты для него можно приобрести у других фирм и на рынке существует спрос на этот продукт. Фирмы могут улучшить свои «kenes», улучшив продукты и рыночные показатели либо за счет отдельных исследований и разработок, либо за счет использования знаний, полученных от других фирм путем совместных мероприятий. Существует два уровня модели. Первый — это промышленный рынок с фирмами, торгующими друг с другом. Второй — это модель, где агенты обучаются путем естественного отбора; фирмы, которые не могут найти клиента, не могут ничего произвести и уходят с рынка, в то время как наиболее конкурентоспособные фирмы выживают, сотрудничают с другими фирмами и производят «дочерние фирмы», которые перенимают лучшие черты «родителей».

Такого рода модели полезны для оптимизации производства и стимулирования экономического и инновационного роста путем создания кластеров.

Управление цепями поставок. Зачастую продукт перед тем, как попасть к пользователю, проходит долгий путь от поставщиков

до дистрибьюторов, что подразумевает наличие множества прочных связей между рядом компаний только для того, чтобы произвести и продать одну единицу продукции. Оптимизация цепи поставок становится все более важным и сложным процессом, так как продукты в настоящее время (особенно в технологически интенсивных секторах) требуют все большего количества компонентов, поставляемых со всего мира. Моделирование цепочек поставок является эффективным способом изучения эффективности управленческой политики, и все большую популярность приобретают мультиагентные модели. Основная причина, по которой мультиагентная модель является наиболее подходящей для этой задачи, заключается в том, что она предоставляет возможность назначить каждому бизнесу, участвующему в цепочке, агента с собственными правилами поведения. Более того, эта модель упрощает имитацию передачи информации и продуктов по цепочке.

Примером использования такого метода может быть изучение влияния обмена информацией на различные цепочки поставок сборных изделий, аналогичных типичному продукту электронной и компьютерной промышленности. Для различных сборочных цепочек поставок характерно небольшое число поставщиков, производящих общие подкомпоненты (например, транзисторы или контроллеры), которые используются для производства ряда типовых продуктов (например, микросхемы, ЦП, жесткие диски), которые затем применяются в точках распространения для сборки специализированных продуктов, таких как персональные компьютеры. Модель учитывает три различные политики выполнения заказов:

1. *Сделать на заказ*, когда производство инициируется заказами клиента.
2. *Собрать на заказ*, когда компоненты собираются и хранятся на складе, и только окончательная сборка инициируется по заказу.
3. *Сделать на запас*, когда производство начинается, если уровень запасов снижается ниже определенного порогового значения.

Кроме того, авторы исследовали, как различное количество обмена информацией между агентами влияет на модель. Результаты показали, что наиболее эффективной моделью является подход «Собрать на заказ», наряду с обменом информацией о спросе и пред-

ложении между организациями, связанными в цепочке поставок. Как отмечали авторы, их результаты подтверждают идею о том, что складские запасы могут быть успешно заменены информацией. Другими словами, если производитель имеет достоверную и достаточную информацию, неопределенность в отношении спроса может быть уменьшена, что позволит оптимизировать имеющиеся запасы.

Модели рынка электроэнергии. В ряде развитых стран электроснабжение приватизировано. Обычно рынок электроэнергии состоит из двух или трех крупных поставщиков электроэнергии, продающих электроэнергию дистрибьюторам, которые затем продают ее потребителям. Поскольку эта структура отличается от традиционной государственной электроэнергетики, она создала поле для исследований с использованием АВМ. Обычно эти модели нацелены на понимание последствий регулирования рынка, изменений в количестве и типе поставщиков и покупателей и учет предполагаемых изменений политики (например, снижение вероятности отключений электроэнергии или снижения вредного воздействия производства на окружающую среду).

Как правило, в этом типе моделей агентами выступают снабженческие компании, которые предлагают рынку обеспечить его определенным количеством электроэнергии по определенной цене на определенный период. Это похоже на реальный рынок, где принимаются лучшие из таких предложений. Однако ситуация на рынке осложняется несколькими факторами. Во-первых, постоянно меняется спрос на электроэнергию, что усложняет процесс установления поставщиками желаемой цены. Второй фактор, связанный с первым, заключается в том, что стоимость выработки электроэнергии нелинейна. Только пиковый спрос, длящийся несколько часов, может потребовать полной мощности производства электроэнергии, в другое время нет смысла производить электроэнергию на всех станциях или на полной мощности.

Существует ряд полезных результатов, которые можно получить, запустив моделирование. Например, можно определить, каково оптимальное количество поставщиков для рынка с заданными параметрами, или при каких условиях цена электроэнергии сводится к ее предельной стоимости генерации, или как различные

рыночные структуры влияют на производителей. В последнее время этот тип моделей был усовершенствован функцией, позволяющей агентам изучать торговые стратегии, чтобы использовать те стратегии, успешность которых была доказана.

Партисипативное (сопутствующее) моделирование. Еще одной удивительной областью, в которой можно применять АВМ, является рациональное использование дефицитных природных ресурсов, таких как вода или леса. Этот подход предполагает построение мультиагентной системы в тесной связи с информаторами, отобранными из числа людей, проживающих в проблемной зоне, и гораздо более практико-ориентирован, чем другие подходы. Участники обычно опрашиваются на предмет проблемной ситуации, а затем учувствуют в специально подобранной ролевой игре (RPG) для сбора данных. После того как все необходимые данные собраны, исследователи строят компьютерную модель, которая используется в качестве учебного пособия или в качестве переговорной поддержки, позволяя ответить на вопросы «что, если...?» о возможных решениях проблемы ограниченных природных ресурсов.

В качестве примера можно привести работу О. Барретау и др. [10], в которой описывается, как использовалось партисипативное моделирование, чтобы понять, почему ирригационная схема в долине реки Сенегал не была успешной в целом. Исследователи создали RPG и многоагентную АВМ под названием «*SHADOC*» для воссоздания взаимодействия между различными заинтересованными сторонами, участвующими в принятии решений о распределении воды на пахотных участках орошаемой территории. Сначала была разработана АВМ, а затем с целью проверки модели и удобства использования среди участников ее основные элементы были преобразованы в ролевую игру, где игроки были эквивалентны агентам. Авторы заявляют, что их подход «усиливает дискуссию среди участников процесса», и в качестве успеха отмечают, что «проблемы, возникающие в этой области и известные каждому человеку отдельно, превратились в общие знания».

Модели, учитывающие историческую специфику (History — Friendly models). Другим типом моделей, которые подпадают под широкий класс АВМ, но имеют несколько существенных отличий

от других типов (и, следовательно, нуждаются в более подробном обзоре), являются «*History — Friendly*»-модели. Они, как и многие другие модели, могут имитировать сложные системы с точки зрения социальных взаимодействий, которые заранее не закодированы. Эти взаимодействия позволяют изучать весь процесс, опираясь на его части. Как указано в Garavaglia (2010), *History — Friendly*-модели являются автономными моделями моделирования, точно определенными и формально структурированными, основываются на определении «взаимодействующих объектов», причем объекты могут представлять как агентов, так и среду, определяемые их характеристиками. Этот тип моделей направлен не на прогнозирование, а на понимание и объяснение поведения определенной отрасли. Он представляет собой описание рутины и поведения экономических агентов в отношении эволюционных механизмов отрасли, таких как появление новых технологий или гибель компаний. Эти модели обычно отвечают на вопросы типа «что могло бы произойти, если...?» и «почему именно это произошло?». Традиционно выделяются три основные точки интереса для *History — Friendly*-моделей:

1. Объяснение современных явлений с точки зрения истории.
2. Изучение правил и условий, которые заставляют событие происходить при определенных ограничениях.
3. Объяснение условий «типичной» истории, которая уже произошла.

Более ранние версии эволюционных моделей обычно имели слишком простую структуру и их эмпирическую основу составляли очень широкие явления. *History — Friendly*-модели обладают более сложной структурой, а их эмпирическая основа имеет ограничения. Эти модели пытаются объединить, с одной стороны, формализованные отношения агентов, с другой — исторические события. Они обычно рассматривают конкретную отрасль или рынок, в то время как другие эволюционные модели могут быть применены к широкому кругу отраслей.

Следует иметь в виду, что целью такого подхода является не реконструкция социально-экономических свойств, а исследование исторических стилизованных фактов. Структура исторически чистых моделей представляет собой, по существу, лабораторную

конструкцию искусственной промышленности с искусственными агентами и рыночными средами, которая сделана для того, чтобы объяснить определенные явления. Это также позволяет назвать *History — Friendly*-модели отраслевыми, поскольку явления исследуются только в конкретной отрасли. Анализ в этом случае состоит в основном из тестирования поведения модели, помогающего понять, какие факторы и фундаментальные процессы заставляют модель действовать так, как она действует. Чтобы это тестирование было успешным и давало надежные результаты, исследователь должен генерировать, а затем сэмплировать данные так, чтобы они были хорошим приближением реальных явлений и процессов. Однако *History — Friendly*-моделирование не направлено на воспроизведение результатов, которые эмпирически полностью соответствуют количественным конкретным значениям имеющихся исторических данных, а скорее должно быть близко к реальным данным в качественном смысле. Прежде всего, модель должна быть логичной и последовательной для достижения своей главной цели. Однако если результаты модели близки к реальности, это обычно считается хорошим показателем. Таким образом, *History — Friendly*-модели исследуют аспекты выбранных явлений как с количественной, так и с качественной точки зрения.

History — Friendly-подход широко использует так называемый контрфактуальный анализ (также известный как *if — worlds*, *allohistory*, *cliometrics*, *uchronia*). Модели, учитывающие историческую специфику, могут даже рассматриваться в качестве нормативного инструмента для изучения того, какое воздействие оказывают различные социально-экономические условия и государственная промышленная политика на лиц, принимающих решения. В целом этот анализ предполагает «изменение» предшествующего события в целях изучения его влияния на последующую ситуацию, тем самым оценивается, как некоторые изменения в прошлом изменят динамику будущих событий. Этот анализ применяется достаточно давно и обычно используется для оценки последствий тех или иных политических решений. Хотя дискуссия об эффективности контрфактуального анализа все еще открыта (в свете того, что его не всегда легко и целесообразно проводить), очевидно, что «хоро-

шие» контрфактуальные исследования являются информативными, полезными и позволяют избежать некоторых социальных предубеждений (например, культурных). Для того чтобы контрфактуальное исследование было хорошим, оно должно быть представлено в простой и понятной форме, с подробными рассуждениями, почему рассматриваемые изменения правдоподобны и почему они не произошли изначально.

Что касается *History — Friendly*-моделей, то контрфактуальные исследования могут использоваться как процедура тестирования и как инструмент исследования. Такой подход может быть хорошей проверкой логичности модели и надежности результатов. Данный анализ можно рассматривать как реконструкцию альтернативных исторических сценариев, которая позволяет рассуждать о путях эволюции отрасли.

Считается, что этот подход, специализированный на одной отрасли и сфокусированный на качественных показателях, в значительной степени опирающийся на историю, представляет собой «падение одного конкретного листа с дерева». Такие модели критикуются именно потому, что они моделируют «типичную» историю, которая может не существовать в реальности или быть просто совпадением. В противовес этому следует отметить, что исторические события, используемые в моделях, обычно являются результатом причинно-следственных связей и систематических объяснений, а не случайного возникновения. Это также может быть причиной того, почему *History — Friendly*-модели должны быть тщательно протестированы и откалиброваны, иногда даже с использованием экстремальных сценариев, где результаты легко предсказуемы.

Наряду с общим взглядом на *History — Friendly*-модели, о котором мы упомянули, есть попытки разработать более общие модели, объясняющие «падение многих листьев со многих деревьев». Возможно, более глубокое понимание основных процессов различных отраслей позволит найти среди них общие черты и правила, которые сделают возможным обобщение. Тем не менее при нынешнем состоянии исследований, посвященных *History — Friendly*-моделям, это не кажется легким для выполнения и оставляет простор для дальнейших исследований.

Прежде чем перейти к примерам, давайте сначала представим основные моменты, которые должны быть освещены при описании *History — Friendly*-модели (и многих других ABM):

1. Ясное объяснение используемой среды. Это, с одной стороны, позволяет лучше понять модель, а с другой стороны, помогает исследователю получить информативные результаты, что придаст модели бóльшую достоверность.

2. Обзор теоретических основ, которые могут быть использованы при анализе уже существующих исследований и теорий. Этот момент служит цели «укрепления» модели и рассматриваемых явлений. Крайне важно дать четкое объяснение того, как эта теоретическая основа включена в структуру модели. Помимо всего прочего, здесь исследователь сталкивается с компромиссом между описательной точностью модели и объяснительной силой.

3. Схема наиболее релевантных характеристик модели. Это, очевидно, необходимо делать для понимания структуры и принципов функционирования модели.

4. Обзор результатов, который включает в себя результаты калибровки, заключающейся в многократном изменении параметров и методов оценки. Если используется контрфактуальный анализ, то он должен быть описан в этой части. Каждый результат должен быть максимально четким, чтобы его можно было использовать в дальнейших исследованиях. Дальнейшие перспективы модели также могут быть упомянуты в этом разделе.

Рассмотрим теперь примеры практического использования таких моделей. Можно отметить исследование Ф. Малерба и др. [11], посвященное эволюции компьютерной промышленности. В статье анализируется ряд ключевых событий компьютерного сектора, например, становление IBM как ведущей фирмы мейнфрейм-рынка, внедрение микропроцессоров, приведшее к росту сегмента рынка персональных компьютеров. Первоначальная модель демонстрирует ключевые факторы в развитии отрасли, такие как сроки внедрения новой технологии, большая фокусировка потребителей на рынке мейнфреймов по сравнению с рынком ПК и более высокая ценовая чувствительность потребителей на рынке ПК.

В дальнейшем авторы усовершенствовали первоначальную модель для рассмотрения эффективности экономической политики в сложных рыночных условиях (имеется в виду компьютерная промышленность с динамично растущей доходностью и высоким уровнем локализации). Это дополнение усложнило модель, поскольку в ней учитывались не только компании и потребители, но и другие государственные или частные учреждения. Безусловно, такое дополнение сделало модель более полной и приблизило ее к реальности.

Эти модели могут также использоваться для изучения последствий новых инновационных подходов на рынке, например, для изучения перехода фармацевтической промышленности от эпохи случайного скрининга к эпохе молекулярной биологии. Данный процесс включал изменения в подходах к разработке новых лекарств, а также изменения в комплектации и структуре рынка. Исследование показало, что с учетом характера процессов разработки лекарственных средств, основных особенностей процесса конкуренции, низкой степени кумулятивности и фрагментарности рынков уровень общей концентрации отрасли вскоре станет достаточно низким. При этом появление биотехнологических фирм не означает вытеснение основных игроков.

В заключение следует также рассмотреть перспективы агентного моделирования и возможные пути его развития. Например, К. Френкен [7] указывает четыре возможных направления развития инновационных моделей (то же самое можно сказать и о всех других АВМ):

1. Применение базовых моделей по темам, еще не исследованным в рамках данного подхода, таким как экономическая география, эволюционная макроэкономика, CSR и многим другим.

2. Объединение основных подходов. Например, объединение моделей цепочек поставок с промышленными сетями позволяет получить интересную информацию о том, как цепочки поставок влияют на кластеризацию компаний, и наоборот.

3. Рекомбинация АВМ с более ранними эволюционными моделями динамики.

4. Эмпирическая проверка моделей. Это можно сделать несколькими способами. В некоторых приложениях гипотезы могут

быть выведены дедуктивно в качестве результатов модели. Хотя в большинстве случаев эмпирическая валидация АВМ основывается на иной методологии, основанной на способности модели воспроизводить стилизованные факты.

Подводя итог, следует отметить, что данный подход не охватывает все существующие реализации агентного моделирования, а лишь показывает, что данный метод моделирования предоставляет исследователям широкие возможности для анализа и тестирования социально-экономической и экологической среды.

ЗАКЛЮЧЕНИЕ

В данном учебном пособии рассматриваются несколько важных аспектов эконометрического анализа и методов Монте-Карло (ММК), позволяющих решать широкий круг задач.

Исходя из вышеизложенного, эконометрический анализ очень важен в современных экономических исследованиях, и никто не может представить хороший экономический документ или исследовательский проект без эмпирики. Эконометрика используется во всех прикладных областях экономики для проверки экономических теорий, для информирования государственных и частных политиков и для прогнозирования экономических временных рядов. Иногда эконометрическая модель основывается на формальной экономической модели, в других случаях — на неформальных экономических рассуждениях и интуиции. Целью любого эконометрического анализа является оценка параметров модели и проверка гипотез об этих параметрах; значения и знаки параметров определяют обоснованность экономической теории и влияние определенных политик.

Каждое экономическое исследование проводится на базе трех основных типов данных: пространственных данных, временных рядов и панельных данных. В главах учебного пособия дается подробный анализ преимуществ, проблем и методов оценки для каждого типа.

Пространственные и временные ряды, объединенные пространственные и панельные данные являются наиболее распространенными типами структур данных, которые используются в прикладной эконометрике. Наборы данных, содержащие понижение размерности (временные ряды и панельные данные), требуют специальной корректировки из-за корреляции во времени большинства экономических временных рядов. Существуют и другие проблемы, такие как тренды и сезонность, возникающие при анализе данных временных рядов, но не пространственных данных.

Каждый исследователь должен учитывать все главные особенности типов данных, базы данных, целей исследований и методов оценки, так как чем правильнее сделан выбор, тем точнее будут результаты и прогнозы.

Классические эконометрические методы не всегда применимы, поэтому в данном пособии мы подчеркиваем необходимость ознакомления современных исследователей с методами Монте-Карло. Основное преимущество ММК заключается в том, что они позволяют вместо упрощения моделей или использования узкоспециализированных подходов применять более общие алгоритмы и при этом получать надежные результаты. Учитывая обычную простоту этих алгоритмов, методы Монте-Карло кажутся отличным инструментом, который может обеспечить неожиданные результаты в сложных задачах.

Мы рассмотрели несколько основных концепций и подходов, использующих идеи методов Монте-Карло. Кроме того, мы показали, что ММК могут применяться к различным проблемам.

Все описанное выше действительно показывает, что методы Монте-Карло могут быть использованы в широком спектре теоретических исследований и прикладных проблем, от физики и биологии до социальных исследований и теорий мнений. Этот диапазон также включает в себя экономику и эконометрику, где методы Монте-Карло могут быть использованы для предоставления информации о сложных системах, которые включают в себя множество экономических агентов и/или многие существенные факторы, влияющие на систему.

Мы также упомянули несколько проблем, с которыми сталкиваются подходы Монте-Карло. Основная из них заключается в том, что исследователи во многих областях не уверены, когда и как

использовать ММК должным образом, и скорее полагаются на традиционные подходы, иногда излишне упрощая модель. Существенным недостатком, особенно в эконометрике, считается «случайность» подходов. Обе эти проблемы можно было бы решить в основном популяризацией ММК и разработкой простых и понятных способов их реализации. Мы также предлагаем использовать методы Монте-Карло в качестве инструмента для оценки результатов традиционных подходов, что облегчит выбор основного метода оценки. Другая важная проблема заключается в том, что с развитием компьютерных технологий, которые когда-то сделали методы Монте-Карло жизнеспособными, появились более продвинутые и эффективные методы оценивания, а рост вычислительной мощности персональных компьютеров позволяет решать больше задач традиционным способом. Эта проблема, очевидно, не может быть решена полностью, но разработка новых подходов Монте-Карло и модификация старых могли бы сохранить эффективность и конкурентоспособность ММК.

Есть несколько основных путей дальнейшего изучения методов Монте-Карло. Одним из них является рассмотрение теоретических рамок более сложных подходов Монте-Карло, например, марковской цепи Монте-Карло. Они больше подходят для сложных исследований и реализаций, которые включают в себя многомерные распределения. Это даст более детальное представление о природе подходов Монте-Карло, позволяя понять, где и когда они могут быть применены. Другой путь — исследование успешных реализаций метода Монте-Карло. Мы представили некоторые из них, которые более или менее связаны с основными описанными подходами, но есть огромное количество других исследований, которые дают представление о том, как методы Монте-Карло могут быть применены для решения широкого спектра кардинально различающихся проблем.

В заключение следует отметить, что методы Монте-Карло уже зарекомендовали себя как эффективный способ решения теоретических и прикладных задач и использовались в многочисленных исследованиях. В то же время они постоянно развиваются и, мы надеемся, будут еще чаще использоваться в будущем.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Wooldridge J. M.* Introductory Econometrics: a modern approach. Michigan State University, 2012. 912 p.
2. *Verbeek M.* A guide to modern econometrics. England : John Wiley & Sons Ltd., 2004. 446 p.
3. *Shonkwiler R. W., Mendivil F.* Explorations in Monte Carlo Methods. N. Y. : Springer, 2009. 243 p.
4. *Steyvers M.* Computational Statistics with Matlab // UCI Cognitive Science Experiments : [сайт]. 2011. URL: <http://psiexp.ss.uci.edu/research/teachingP205C/205C.pdf> (дата обращения: 28.08.2018).
5. *Gilli M., Winker P.* Review of Heuristic Optimization Methods in Econometrics // Working Papers 001. COMISEF, 2008.
6. *Blüschke D., Blüschke-Nikolaeva V., Savin I.* New insights into optimal control of nonlinear dynamic econometric models: application of a heuristic approach // Journal of Economic Dynamics and Control. 2013. № 37(4). P. 821–837.
7. *Frenken K.* Technological Innovation and Complexity Theory // Econ. Innov. New Techn. 2006. Vol. 15(2). P. 137–155.
8. *Edmonds B., Moss S.* From KISS to KIDS — an ‘anti-simplistic’ modelling approach // CPM Report. 2004. № 04–132. P. 15.
9. *Gilbert N.* Agent-based models. Sage Publications Inc., 2007. 122 p.

10. *Barreteau O., Bousquet F., Attonaty J.-M.* Role-Playing Games for Opening the Black Box of Multi-Agent Systems: Method and Lessons of its Application to Senegal River Valley Irrigated Systems // *Journal of Artificial Societies and Social Simulation*. 2001. Vol. 4, № 2. P. 13.

11. *Malerba F., Nelson R. R., Orsenigo L., Winter S. G.* History-friendly models of industry evolution: the computer industry // *Industrial and Corporate Change*. 1999. Vol. 8. P. 3–40.

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

Aarts E. Simulated Annealing and Boltzmann Machines / E. Aarts, J. Korst. Chichester : Wiley and Sons, 1989.

Acosta-González E. Model selection via genetic algorithms illustrated with cross-country growth data / E. Acosta-González, F. Fernández-Rodríguez // Empirical Economics. 2007. № 33. P. 313–337.

Adanu K. Optimizing the GARCH model — an application of two global and two local search methods / K. Adanu // Computational Economics. 2006. № 28. P. 277–290.

Ahn S. K. Estimation for partially nonstationary multivariate autoregressive models / S. K. Ahn, G. C. Reinsel // Journal of the American Statistical Association. 1990. № 85(411). P. 813–823.

Alcock J. A genetic estimation algorithm for parameters of stochastic ordinary differential equations / J. Alcock, K. Burrage // Computational Statistics and Data Analysis. 2004. № 47(2). P. 255–275.

Allen P. M. Cities and Regions as Self-Organizing Systems: Models of Complexity / P. M. Allen. Amsterdam : Gordon & Breach, 1997.

Althöfer I. On the convergence of threshold accepting / I. Althöfer, K.-U. Koschnick // Applied Mathematics and Optimization. 1991. № 24. P. 183–195.

Ashlock D. Evolutionary Computation for Modeling and Optimization / D. Ashlock. N. Y. : Springer-Verlag, 2006.

Atkinson A. C. Simulated annealing for the detection of multiple outliers using least squares and least median of squares fitting / A. C. Atkinson, S. W. Weisberg // *Directions in robust statistics and diagnostics. Pt. 1* (W. A. Stahel, S. W. Weisberg, eds.). N. Y. : Springer-Verlag, 1991.

Axelrod R. The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration / R. Axelrod. Princeton ; N. J. : Princeton University Press, 1997.

Bäck T. Handbook of Evolutionary Computation / T. Bäck, D. Fogel, Z. Michalewicz. Oxford, 1997.

Bagnall A. J. A Multiagent Model of the UK Market in Electricity Generation. Evolutionary Computation / A. J. Bagnall, G. D. Smith // *IEEE Transactions*. 2005. № 9 (5). P. 522–536.

Baragona R. Fitting piecewise linear threshold autoregressive models by means of genetic algorithms / R. Baragona, F. Battaglia, D. Cucina // *Computational Statistics & Data Analysis*. 2004. № 47. P. 277–295.

Baragona R. Genetic algorithms for the identification of additive and innovation outliers in time series / R. Baragona, F. Battaglia, C. Calzini // *Computational Statistics & Data Analysis*. 2001. № 37(1). P. 1–12.

Barreteau O. Role-Playing Games for Opening the Black Box of Multi-Agent Systems: Method and Lessons of its Application to Senegal River Valley Irrigated Systems / O. Barreteau, F. Bousquet, J.-M. Attonaty [Electronic resource] // *Journal of Artificial Societies and Social Simulation*. 2001. Vol. 4, № 2. URL: <http://jasss.soc.surrey.ac.uk/4/2/5.html>.

Batten D. NEMSIM: Finding Ways to Reduce Greenhouse Gas Emissions using Multi-Agent Electricity Modelling / D. Batten, G. Grozev // Chapter 11 of *Complex Science for a Complex World: Exploring Human Ecosystems with Agents* (P. Pascal and D. Batten, eds.). ANU Press, 2006. P. 227–254.

Bauer D. Estimating cointegrated systems using subspace algorithms / D. Bauer, M. Wagner // *Journal of Econometrics*. 2002. № 111. P. 47–84.

Beretvas S. N. Cross-classified random effects models / S. N. Beretvas // *Multilevel modeling of educational data*. Charlotte, NC Information Age Publishing, 2008.

Birattari M. Classification of metaheuristics and design of experiments for the analysis of components / M. Birattari, L. Paquete, T. Stützle, K. Varrentrap // Technical Report AIDA-2001–05. Intellektik, Technische Universität Darmstadt, 2001.

Blüschke D. New insights into optimal control of nonlinear dynamic econometric models: application of a heuristic approach / D. Blüschke, V. Blüschke-Nikolaeva, I. Savin // *Journal of Economic Dynamics and Control*. 2013. № 37(4). P. 821–837

Bock F. An algorithm for solving «traveling salesman» and related network optimization problems / F. Bock. St. Louis : 14th ORSA meeting, 1958.

Boden M. A. Computer models of mind / M. A. Boden. Cambridge University Press, 1988.

Bratley P. A Guide to Simulation / P. Bratley, B. Fox, L. Scrage. N. Y. : Springer-Verlag, 1983.

Brooks S. P. Classical model selection via simulated annealing / S. P. Brooks, N. Friel, R. King // *Journal of the Royal Statistical Society Series*. 2003. № B65. P. 503–520.

Brooks S. P. Optimization using simulated annealing / S. P. Brooks, B. J. T. Morgan // *The Statistician*. 1995. № 44(2). P. 241–257.

Brown B. W. Statistics: A Biomedical Introduction / B. W. Brown, M. Hollander. N. Y. : Wiley, 1977.

Carley K. M. On Generating Hypotheses Using Computer Simulations / K. M. Carley // *Systems Engineering*. 1999. Vol. 2, iss. 2. P. 69–77.

Chipman J. S. Optimal aggregation of linear time series models / J. S. Chipman, P. Winker // *Computational Statistics and Data Analysis*. 2005. № 49 (2). P. 311–331.

Chung K. L. A course in probability theory / K. L. Chung. N. Y. : Academic Press, 1974.

Cohen M. D. A Garbage Can Model of Organizational Choice / M. D. Cohen, J. G. March, J. P. Olsen // *Administrative Science Quarterly*. 1972. № 17 (1). P. 1–25.

Colnari A. An investigation of some properties of an ant algorithm / A. Colnari, M. Dorigo, V. Manniezzo // *Parallel problem solving from nature*. 1992. Vol. 2. (R. Männer, B. Manderick, ed.). Amsterdam : North-Holland. P. 509–520.

Colnari A. Distributed optimization by ant colonies / A. Colnari, M. Dorigo, V. Manniezzo // *Proceedings of the First European Conference on Artificial Life (ECAL-91)* (F. J. Varela, P. Bourguine, ed.). Cambridge MA : The MIT Press, 1992. P. 134–142.

Croes A. A Method for solving traveling salesman problem / A. Croes // *Operational Research*. 1958. № 5. P. 798–812.

Deffuant G. How Can Extremism Prevail? A Study Based on the Relative Agreement Interaction Model [Electronic resource] / G. Deffuant, F. Amblard, G. Weisbuch, T. Faure // Journal of Artificial Societies and Social Simulation. 2002. Vol. 5, № 4. URL: <http://jasss.soc.surrey.ac.uk/5/4/1.html>.

Dorsey B. Genetic algorithms for estimation problems with multiple optima, nondifferentiability and other irregular features / B. Dorsey, W. J. Mayer // Journal of Business and Economic Statistics. 1995. № 13. P. 53–66.

Dougherty C. Introduction to econometrics / C. Dougherty. Oxford University Press, 2000. 335 p.

Dueck G. Threshold accepting: A general purpose algorithm appearing superior to simulated annealing / G. Dueck, T. Scheuer // Journal of Computational Physics. 1990. № 90. P. 161–175.

Eberhart R. C. A new optimizer using particle swarm theory / R. C. Eberhart, J. Kennedy // Proceedings of the Sixth International Symposium on Micromachine and Human Science. Nagoya, Japan, 1995. P. 39–43.

Edmonds B. From KISS to KIDS — an ‘anti-simplistic’ modelling approach [Electronic resource] / B. Edmonds, S. Moss // CPM Report. 2004. № 04–132. URL: <http://cfpm.org/cpmreps.html>.

Enders W. Applied Econometric Time Series. 2nd ed. / W. Enders. N.Y. : John Wiley and Sons, 2004. 544 p.

Frenken K. Technological Innovation and Complexity Theory / K. Frenken // Econ. Innov. New Techn. 2006. Vol. 15(2). P. 137–155.

Gan L. A test for global maximum / L. Gan, J. Jiang // Journal of the American Statistical Association. 1999. № 94(447). P. 847–854.

Garavaglia C. Modelling Industrial Dynamics with «History-Friendly» Simulations / C. Garavaglia // Structural Change and Economic Dynamics. 2010. № 21. P. 258–275.

Gilbert N. Agent-based models / N. Gilbert. Sage Publications Inc., 2007.

Gilbert N. Innovation networks — A simulation approach / N. Gilbert, A. Pyka, P. Ahrweiler // Journal of Artificial Societies and Social Simulation. 2001. Vol. 4, № 3. U131–U150.

Gilbert N. Varieties of Emergence / N. Gilbert // University of Chicago and Argonne National Laboratory Sociology: Agent 2002 Conference: Social agents: ecology, exchange, and evolution. Chicago, 2002. P. 41–56.

Gillman L. The Car and the Goats / L. Gillman // American Mathematical Monthly. 1992. Vol. 99. P. 3–7.

Givens G. Computational Statistics / G. Givens, J. Hoeting. N. Jersey : Wiley-Interscience. Hoboken, 2005.

Glover F. Tabu Search / F. Glover, M. Laguna. Boston, MA : Kluwer Academic Publishers, 1997.

Goffe W.L. Global optimization of statistical functions with simulated annealing / W.L. Goffe, G.D. Ferrier, J. Rogers // Journal of Econometrics. 1994. № 60(1–2). P. 65–99.

Goldberg D.E. Genetic Algorithms in Search. Optimization, and Machine Learning / D.E. Goldberg. Massachusetts : Addison-Wesley, Reading, 1989.

Goldstein H. Multilevel statistical models / H. Goldstein. L. : Wiley. 2011. 458 p.

Greene W.H. Econometric Analysis. 5th ed. / W.H. Greene. Prentice Hall, 2003. 959 p.

Greene W.H. Econometric Analysis / W.H. Greene. Prentice Hall, 2008. 193 p.

Grimaldi R. Discrete and Combinatorial Mathematics / R. Grimaldi. Addison Wesley, 2003.

Hajek B. Cooling schedules for optimal annealing / B. Hajek // Mathematics of Operations Research. 1988. Vol. 13, № 2. P. 311–329.

Hammersley J.M. Monte Carlo Methods / J.M. Hammersley, D.C. Handcomb. N. Y. : Wiley, 1964.

Hawkins D.S. Determining the number of components in mixtures of linear models / D.S. Hawkins, D.M. Allen, A.J. Stromberg // Computational Statistics & Data Analysis. 2001. № 38(1). P. 15–48.

Holland J.H. Adaptation in Natural and Artificial Systems / J.H. Holland. Ann Arbor, MI : The University of Michigan Press, 1975.

Hull J.C. Options, Futures, and Other Derivatives. 5rd ed. / J.C. Hull. N. Jersey : Prentice Hall, Upper Saddle River, 2003.

Hüsler J. On optimization and extreme value theory / J. Hüsler, P. Cruz, A. Hall, C.M. Fonseca // Methodology and Computing in Applied Probability. 2003. № 5. P. 183–195.

Izquierdo S.S. Market Failure Caused by Quality Uncertainty / S.S. Izquierdo, L.R. Izquierdo, J.M. Galán, C. Hernández // Lecture Notes in Economics and Mathematical Systems. 2006. Vol. 564. P. 203–213.

Jacobson S.H. Global optimization performance measures for generalized hill climbing algorithms / S.H. Jacobson, E. Yücesan // Journal of Global Optimization. 2004. № 29. P. 173–190.

Jacobson S. H. Visiting near-optimal solutions using local search algorithms / S. H. Jacobson, S. N. Hall, L. A. McLay // COMPSTAT 2006: Proceedings in Computational Statistics (A. Rizzi and M. Vichi, ed.). Physica. Heidelberg, 2006. P. 471–481.

Janssen M. An Integrated Approach to Simulating Behavioural Processes: a Case Study of the Lock-in of Consumption Patterns [Electronic resource] / M. Janssen, W. Jager // Journal of Artificial Societies and Social Simulation. 1999. Vol. 2, № 2. URL: <http://jasss.soc.surrey.ac.uk/2/2/2.html>.

Jerrell M. E. Global optimization of econometric functions / M. E. Jerrell, W. A. Campione // Journal of Global Optimization. 2001. № 20(3–4). P. 273–295.

Kalos M. H. Monte Carlo Methods / M. H. Kalos, P. A. Whitlock. Wiley-Interscience, 1986.

Kirkpatrick S. Optimization by simulated annealing / S. Kirkpatrick, C. Gelatt, M. Vecchi // Science. 1983. № 220. P. 671–680.

Knuth D. Art of Computer Programming. Vol. 2 : Seminumerical Algorithms. 3rd ed. / D. Knuth. Massachusetts : Addison-Wesley Professional, 1997.

Koesrindartoto D. An Agent-Based Computational Laboratory for Testing the Economic Reliability of Wholesale Power Market Designs / D. Koesrindartoto, J. Sun, L. Tesfatsion // IEEE Power Engineering Society General Meeting. 2005. P. 2818–2823.

Kremer N. S. Econometrics / N. S. Kremer, B. A. Putko. M. : UNITI, 2010. 328 p.

Liu J. Monte Carlo Strategies in Scientific Computing / J. Liu. N. Y. : Springer-Verlag, 2001.

Maddala G. S. Maximum likelihood methods for models of markets in disequilibrium / G. S. Maddala, F. D. Nelson // Econometrica. 1974. № 42(6). P. 303–317.

Maddala G. S. Introduction to Econometrics. 3rd ed. / G. S. Maddala. Chichester : John Wiley and Sons, 2001.

Madras N. Lectures on Monte Carlo Methods / N. Madras // American Mathematical Society. Providence, 2002.

Magnus Y. R. Econometrics. Beginning course / Y. R. Magnus, P. K. Katyshev, A. A. Perecetskii. M. : Delo, 2004. 576 p.

Malerba F. Innovation and market structure in the dynamics of the pharmaceutical industry and biotechnology: towards a History-friendly model / F. Malerba, L. Orsenigo // Industrial and Corporate Change. 2002. № 11. P. 667–703.

Malerba F. Competition and industrial policies in a History-friendly model of the evolution of the computer industry / F. Malerba, R. R. Nelson, L. Orsenigo, S. G. Winter // International Journal of Industrial Organization. 2001. № 19. P. 635–664.

Malerba F. History-friendly models of industry evolution: the computer industry / F. Malerba, R. R. Nelson, L. Orsenigo, S. G. Winter // Industrial and Corporate Change. 1999. № 8. P. 3–40.

Marazzi A. Algorithms, Routines, and S-Functions for Robust Statistics / A. Marazzi. Wadsworth & Brooks/Cole, 1992.

Maringer D. Portfolio Management with Heuristic Optimization / D. Maringer. Dordrecht : Springer, 2005.

Matlab Documentation [Electronic resource]. URL: <http://www.mathworks.com/help/matlab/>.

Matsumoto M. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator / M. Matsumoto, T. Nishimura // ACM Transactions on Modeling and Computer Simulation. 1998. Vol. 8, № 1. P. 3–30.

Metropolis N. The Beginning of the Monte Carlo Method / N. Metropolis // Los Alamos Science Special Issue. 1987. P. 125–130.

Mladenovic N. Variable neighborhood search / N. Mladenovic, P. Hansen // Computers and Operations Research. 1997. № 34. P. 1097–1100.

Moscato P. On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms. Technical report / P. Moscato. Caltech, 1989.

Niederreiter H. Random Number Generation and Quasi-Monte Carlo Methods / H. Niederreiter // CBMS-NSF Regional Conference Series in Applied Mathematics 63. Society for Industrial and Applied Mathematics. Philadelphia, 1992. P. 241.

Ohlmann J. Convergence in Probability of Compressed Annealing / J. Ohlmann, J. Bean, S. Henderson // Mathematics of Operations Research. 2004. № 29, iss. 4. P. 837–860.

Patterson K. D. An Introduction to Applied Econometrics: A Time Series Approach / K. D. Patterson. L. : MacMillan Press, 2000. 795 p.

Reeves C. R. Genetic Algorithms — Principles and Perspectives / C. R. Reeves, J. E. Rowe. Boston : Kluwer, 2003.

Rousseeuw P. J. An Algorithm for Positive-Breakdown Regression Based on Concentration Steps / P. J. Rousseeuw, K. Van Driessen // Data Analysis: Scientific

Modeling and Practical Application (W. Gaul, O. Opitz, M. Schader, ed.). N. Y. : Springer-Verlag. P. 335–346.

Rubenstein R. Y. Simulation and the Monte Carlo Method / R. Y. Rubenstein. N. Y. : John Wiley & Sons, 1981.

Rudolph G. Convergence Properties of Evolutionary Algorithms / G. Rudolph. Hamburg : Kováč, 1997.

Salibian-Barrera M. A Fast Algorithm for S-Regression Estimates / M. Salibian-Barrera, V. J. Yohai // Journal of Computational and Graphical Statistics. 2006. № 15. P. 414–427.

Schelling T. C. Micromotives and Macrobehavior / T. C. Schelling. N. Y. : W. W. Norton, 1978.

Schelling T. C. Dynamic Models of Segregation / T. C. Schelling // Journal of Mathematical Sociology. 1971. № 1. P. 143–186.

Shonkwiler R. W. Explorations in Monte Carlo Methods / R. W. Shonkwiler, F. Mendivil. N. Y. : Springer, 2009.

Simon H. A. Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting / H. A. Simon. N. Y. : John Wiley and Sons, 1957.

Snijders T. A. Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling / T. A. Snijders, R. J. Bosker. L. : Sage, 2011. P. 54–56.

Staszewska A. Representing uncertainty about response paths: The use of heuristic optimization methods / A. Staszewska // Computational Statistics & Data Analysis. 2007. № 52(1). P. 121–132.

Steyvers M. Computational Statistics with Matlab [Electronic resource] / M. Steyvers. 2011. URL: <http://psiexp.ss.uci.edu/research/teachingP205C/205C.pdf>.

Storn R. Differential evolution — A simple and efficient heuristic for global optimization over continuous spaces / R. Storn, K. Price // Journal of Global Optimization. 1997. № 11. P. 341–359.

Strader T. J. Simulation of Order Fulfillment in Divergent Assembly Supply Chains [Electronic resource] / T. J. Strader, F.-R. Lin, M. J. Shaw // Journal of Artificial Societies and Social Simulation. 1998. Vol. 1, № 2. URL: <http://jasss.soc.surrey.ac.uk/1/2/5.html>.

Taillard E. D. Adaptive memory programming: A unified view of metaheuristics / E. D. Taillard, L. M. Gambardella, M. Gendreau, J.-Y. Potvin // European Journal of Operational Research. 2000. № 135. P. 1–16.

Verbic M. Advances in Econometrics: Theory and Applications by Miroslav Verbic / M. Verbic. Publisher : InTech, 2011. 116 p.

Winker P. Application of threshold accepting to the evaluation of the discrepancy of a set of points / P. Winker, K.-T. Fang // SIAM Journal on Numerical Analysis. 1997. № 34. P. 2028–2042.

Winker P. Identification of multivariate AR-models by threshold accepting / P. Winker // Computational Statistics and Data Analysis. 1995. № 20(9). P. 295–307.

Winker P. Optimal lag structure selection in VAR and VEC models / P. Winker, D. Maringer // New Directions in Macromodeling (A. Welfe, Eds.). Amsterdam : Elsevier, 2004. P. 213–234.

Winker P. Optimized multivariate lag structure selection / P. Winker // Computational Economics. 2000. № 16. P. 87–103.

Winker P. Optimization Heuristics in Econometrics: Applications of Threshold Accepting / P. Winker. Chichester : Wiley, 2001.

Wooldridge J. M. Introductory Econometrics: a modern approach / J. M. Wooldridge. Michigan State University, 2013. 912 p.

Учебное издание

Мариев Олег Святославович
Савин Иван Валерьевич
Пушкарев Андрей Александрович
Шорохова Ирина Сергеевна
Жуков Алексей Николаевич
Кисляк Надежда Валерьевна

МЕТОДЫ И МОДЕЛИ ПРИКЛАДНЫХ ИССЛЕДОВАНИЙ

Учебное пособие

Заведующий редакцией *М. А. Овечкина*
Редактор *Н. В. Чапаева*
Корректор *Н. В. Чапаева*
Компьютерная верстка *В. К. Матвеев*

Подписано в печать 30.08.2019 г. Формат 60 × 84 ¹/₁₆.
Бумага офсетная. Цифровая печать. Усл. печ. л. 13,49.
Уч.-изд. л. 11,5. Тираж 40 экз. Заказ 194.

Издательство Уральского университета
Редакционно-издательский отдел ИПЦ УрФУ
620083, Екатеринбург, ул. Тургенева, 4
Тел.: +7 (343) 389-94-79, 350-43-28
E-mail: rio.marina.ovechkina@mail.ru

Отпечатано в Издательско-полиграфическом центре УрФУ
620083, Екатеринбург, ул. Тургенева, 4
Тел.: +7 (343) 358-93-06, 350-58-20, 350-90-13
Факс: +7 (343) 358-93-06
<http://print.urfu.ru>

